DOCUMENT RESUME

ED 275 754 TM 860 671

AUTHOR Perry, Katye M., Ed.; And Others
TITLE State-Mandated Student Assessment Programs: An Examination of the Issues. Proceedings of a Teleconference (October 3, 1985) and a Conference of National and State Educators (Stillwater, Oklahoma, October 4, 1985).
INSTITUTION Oklahoma State Univ., Stillwater. Coll. of Education.
PUB DATE Oct 85
NOTE 122p.
PUB TYPE Collected Works - Conference Proceedings (021) -- Viewpoints (120) -- Reports - General (140)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Achievement Tests; Basic Skills; Educational Economics; Educational Policy; *Educational Testing; Elementary Secondary Education; *Minimum Competency Testing; *State Legislation; State Programs; State Standards; *Testing Problems; *Testing Programs; *Test Use
IDENTIFIERS Florida; Oklahoma

ABSTRACT

These combined proceedings summarize presentations of two conference programs designed to address state-mandated student achievement issues. Dr. Terrel Bell delivered the keynote speech at a teleconference focusing on a broad perspective of competency testing. He discussed testing problems, as well as positive aspects such as comparison, diagnosis and improvement, and public support. A second program presented in the format of a conventional all-day conference, addressed educational, political, philosophical, and psychometric implications of state-mandated assessment programs. Beverly L. Anderson discussed the use of testing for school improvement. David S. Lane, Jr. described the implications of state-level testing on the classroom system. David Yellin also discussed educational implications, such as increased emphasis on basic skills testing which does not put enough emphasis on other educational challenges. Joseph R. Weaver discussed political and financial implications, such as the funding of effective programs. William E. Camp discussed the costs of testing in Oklahoma and other states. Philosphical concerns, such as problems for minority or special students, were raised by Daniel Selakovich. Kay Sather Bull also discussed philosophical implications, such as deciding what should be taught. Janie L. Hall listed a number of practical reservations about minimum competency testing. Psychometric concerns such as scoring, test length, reliability, validity, and standards were reviewed by Dianna Newman. Thomas H. Fisher described Florida's testing experience, and Carolyn W. Smith described Oklahoma's program. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

An Examination of the Issues

ED275754

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. M. Perry

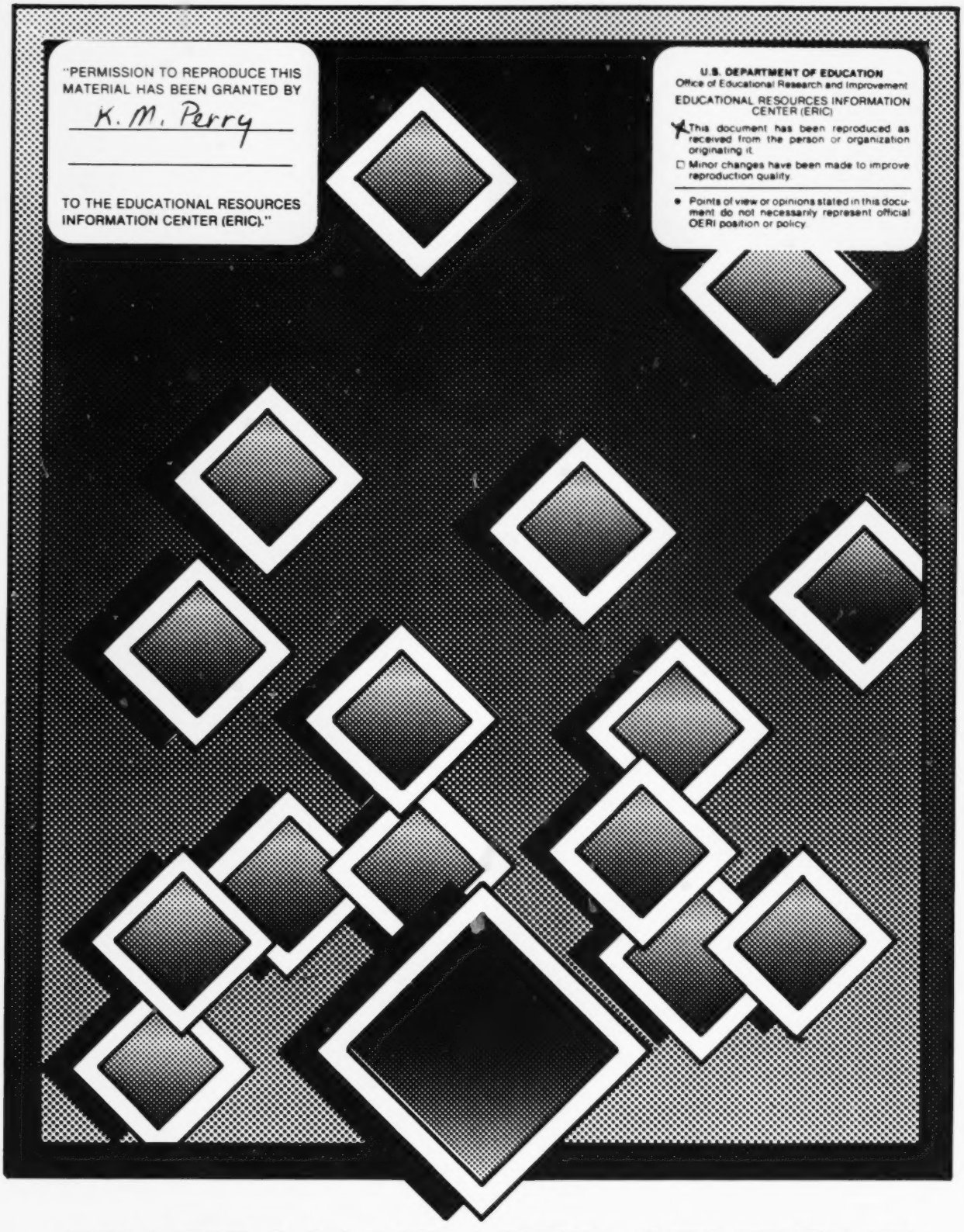
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

★ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

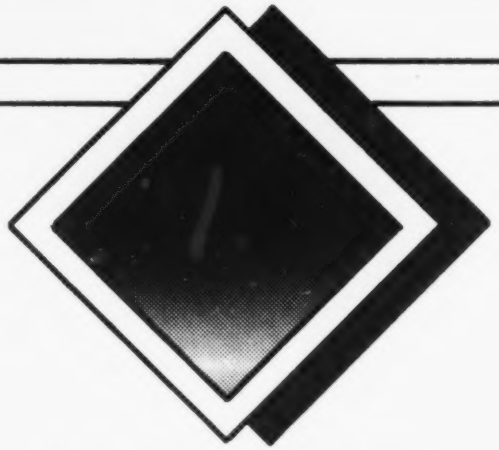
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.



STATE-MANDATED STUDENT ASSESSMENT PROGRAMS

TM 860 671

STATE-MANDATED STUDENT ASSESSMENT PROGRAMS



The Proceedings of
the Teleconference held on
October 3, 1985
featuring Dr. Terrel H. Bell
former U.S. Secretary of Education.

AND

An Examination of the Issues
by National and State Educators
at the Conference held on
October 4, 1985.

EDITED BY:

Katye M. Perry, Assistant Professor
N. Jo Campbell, Associate Professor
David S. Lane, Assistant Professor

Department of Applied Behavioral Studies
College of Education
OKLAHOMA STATE UNIVERSITY

OSU

Oklahoma State University

OFFICE OF THE DEAN
COLLEGE OF EDUCATION

STILLWATER, OKLAHOMA 74078
GUNDERSEN HALL 103
(405) 624-6346

October 7, 1985

Ladies and Gentlemen:

These proceedings represent a summary of the first in a series of five staff development teleconferences produced by the Oklahoma State University College of Education during the 1985-86 academic year. The first of the series is unique in that it is comprised of a teleconference followed by a statewide campus conference. We are particularly excited about these proceedings--first, because the teleconference proceedings feature Dr. Terrel Bell, who, in his tenure as Secretary of Education, focused the nation's attention on many of the issues critical to our search for quality in education, and second, because the campus conference proceedings expand upon this timely topic which is evoking a great deal of thoughtful discussion among educational theorists and practitioners, state legislators, parents, and concerned citizens alike.

State-mandated student assessment is an issue in the public forum. Therefore, it is our intent in these proceedings to spotlight some of the topics of vital interest to educators by featuring the views of many experts who can help us explore some of the more important dimensions of these issues.

Thanks to all presenters and participants. Results, when communicated, can contribute importantly to the continuation of the education movement in Oklahoma.

Sincerely,

Donald W. Robinson, Dean
College of Education
Oklahoma State University

CENTENNIAL DECADE
1980 • 1990

OSU

Oklahoma State University

APPLIED BEHAVIORAL STUDIES IN EDUCATION
COLLEGE OF EDUCATION

STILLWATER, OKLAHOMA 74078
116 NORTH MURRAY HALL
(405) 624-6040

October 7, 1985

State-mandated student achievement issues remain prominent in education today as evidenced by increased attention in both educational and public media. The concept of state-mandated testing has been met with acceptance, contradiction, and controversy.

In an effort to address some of the many issues that surround this concept, two conference program formats were planned. The first was a national teleconference focusing on a broad perspective of competency testing and featured Dr. Terrel H. Bell, Former U.S. Secretary of Education. A panel consisting of a public school superintendent, a nationally recognized classroom teacher and a parent involved in state and national education organizations responded from their personal perspectives. Opportunity for additional reactions from viewers were provided.

Serving as a second program format, a conventional all-day conference was planned to address the educational, political, philosophical, and psychometric implications of state-mandated assessment programs. State legislators, faculty, and psychometric specialists served as workshop leaders. The keynote speakers were Dr. Beverly L. Anderson, Education Commission of the States, and Dr. Thomas H. Fisher, Florida Department of Education.

The contents of these Proceedings represent a summary of both the formal and informal perceptions of the conference leaders and participants. It should be recognized that the contents do not necessarily represent the views of the Department of Applied Behavioral Studies, Oklahoma State University, or the State of Oklahoma. We hope that these contents will provide readers with a more informal view of the issues that surround this timely topic.

We would like to express our sincere appreciation to Dr. Thomas Smith and Constance Lawry for their invaluable assistance with the coordination of both conferences. Thanks are also extended to Dianne Wimberley for her editorial assistance and to Jane Hackett for her word-processing skill in the preparation of this document.

Sincerely,

N. Jo Campbell
Associate Professor

Katye M. Perry
Assistant Professor

David S. Lane
Assistant Professor

CENTENNIAL DECADE 1980 • 1990

PROGRAM COORDINATORS

Dr. N. Jo Campbell, Associate Professor
Department of Applied Behavioral Studies
College of Education
Oklahoma State University
Stillwater, Oklahoma

Dr. Katye M. Perry, Assistant Professor
Department of Applied Behavioral Studies
College of Education
Oklahoma State University
Stillwater, Oklahoma

Dr. David S. Lane, Assistant Professor
Department of Applied Behavioral Studies
College of Education
Oklahoma State University
Stillwater, Oklahoma

TELECONFERENCE PARTICIPANTS

Dr. Terrel H. Bell, Professor of Education
University of Utah
Former U.S. Secretary of Education
Salt Lake City, Utah

Dr. Larry Zenke, Superintendent
Tulsa Public Schools
Tulsa, Oklahoma

Ms. Nancy O'Donnell, Teacher,
Perry Public Schools
(1983-84 1st Runner-Up
National Teacher of the Year)
Perry, Oklahoma

Ms. Karen Leveridge, President
Oklahoma Coalition on Public Education
Oklahoma City, Oklahoma

CONFERENCE SPEAKERS

Dr. Beverly L. Anderson, Associate Director
Education Commission of the States
Denver, Colorado

Dr. David S. Lane, Jr., Assistant Professor
Applied Behavioral Studies in Education
College of Education
Oklahoma State University
Stillwater, Oklahoma

Dr. David Yellin, Associate Professor
Curriculum and Instruction in Education
College of Education
Oklahoma State University
Stillwater, Oklahoma

Dr. Joseph R. Weaver, Director
Teacher Education/Staff Development
Oklahoma State University
Stillwater, Oklahoma

Dr. William E. Camp, Assistant Professor
Educational Administration and Higher Education
College of Education
Oklahoma State University
Stillwater, Oklahoma

Dr. Daniel Selakovich, Professor
Curriculum and Instruction in Education
College of Education
Oklahoma State University
Stillwater, Oklahoma

Dr. Kay Sather Bull, Associate Professor
Applied Behavioral Studies in Education
College of Education
Oklahoma State University
Stillwater, Oklahoma

Ms. Janie L. Hall, Standardized Test Specialist
Oklahoma City Public Schools
Oklahoma City, Oklahoma

CONFERENCE SPEAKERS
(Continued)

Dr. Dianna Newman, Assistant Professor
Applied Behavioral Studies in Education
College of Education
Oklahoma State University
Stillwater, Oklahoma

Dr. Thomas H. Fisher, Director
Assessment, Testing, and Evaluation Section
Florida State Department of Education

Dr. Carolyn W. Smith, Senior Administrative Assistant
for Education
Office of the Governor of the State of Oklahoma
Oklahoma City, Oklahoma

PARTICIPANTS IN A CONVERSATION WITH THE SPEAKERS

Dr. Beverly L. Anderson, Education Commission of the States
Dr. Kay S. Bull, Oklahoma State University
Dr. William E. Camp, Oklahoma State University
Dr. Charles R. (Bob) Davis, Moderator, Oklahoma State University
Dr. Thomas H. Fisher, Florida State Department of Education
Ms. Judy Leach, Oklahoma State Department of Education
Dr. Zoe Leimgruebler, Oklahoma State Department of Education
Dr. Daniel Selakovich, Oklahoma State University
Dr. Joseph R. Weaver, Oklahoma State Department of Education
Dr. David Yellin, Oklahoma State University

TABLE OF CONTENTS

Acknowledgments

Remarks by Dr. Donald W. Robinson, Dean
College of Education, Oklahoma State University. i

Remarks from Program Coordinators: Dr. N. Jo Campbell,
Dr. Katye M. Perry, and Dr. David S. Lane, Jr. iii

Teleconference

Keynote Speaker:
Dr. Terrel Bell. 1
Former U.S. Secretary of Education

Panel Discussion with Dr. Bell. 10

Statewide Conference

Keynote Address: "Using the Psychometric Aspects of
Testing to Promote School Improvement" 21
Dr. Beverly L. Anderson

Concurrent Sessions

·Educational Implications:

"Educational Implications for State-Mandated Student
Assessment on the Classroom System" 33
Dr. David S. Lane, Jr.

"State-Mandated Student Assessment Programs:
Educational Implications" 39
Dr. David Yellin

·Political/Financial Implications:

"State-Mandated Student Assessment Programs:
Political/Financial Implications" 45
Dr. Joseph R. Weaver

"State-Mandated Student Assessment Programs-Finance". . .51
Dr. William E. Camp

·Philosophical Implications:

"State-Mandated Student Assessment: Some
Philosophical Concerns" 55
Dr. Daniel Selakovich

"Draining the Swamp: Philosophical Implications of
State-Mandated Competency-Based Testing". 65
Dr. Kay S. Bull

Psychometric Implications:

"Minimum Competency Testing: Some Practical Concerns". .73
Ms. Janie L. Hall

"Psychometric Issues of State-Mandated
Student Assessement Programs" 83
Dr. Dianna Newman

Keynote Address: "Florida's Experience in Implementing a
Competency-Based Testing Program" 95
Dr. Thomas H. Fisher

A Conversation With the Speakers 105

Keynote Adress: "State-level Involvement in the Competency
Testing of Students". 117
Dr. Carolyn W. Smith

State-Mandated Student Assessment Teleconference

Introduction
by
Dean Donald W. Robinson

Good afternoon. Welcome to the first of a series of five staff development teleconferences to be produced by the Oklahoma State University College of Education for the 1985-1986 academic year. We are particularly excited about this teleconference, first, because it features Dr. Ted Bell, who in his distinguished career as Secretary of Education focused the nation's attention on many of the issues critical to our search for quality in education, and second because it addresses a topic which is evoking a great deal of thoughtful discussion among educators, state legislators, parents, and concerned citizens alike. State-mandated student assessment is an issue in the public forum and it is our intention in this series of staff development teleconferences to spotlight topics of vital interest to educators; to feature experts who can help us explore some of the more important dimensions of these issues, and to address specific questions which are phoned in by those of you receiving the teleconference at sites across the country. The teleconference today as well as those scheduled for later in the year on early education, school law, younger and older learners, and new directions in reading instruction are intended to stimulate continued dialogue within your schools and communities regarding the innovations and challenges confronting those of us in Education today. Again, welcome. We're glad to have you with us. And now to begin today's discussion I would like to turn the program over to Dr. Katye Perry, Assistant Professor in our department of Applied Behavioral Studies and moderator for today's teleconference.

Dr. Perry:

Thank you, Dean Robinson. I wish to welcome you to Oklahoma State University's Telecommunications Center here in Stillwater, Oklahoma, and later to KSLTV in Salt Lake City, Utah. We are very excited about our program today for it addresses a topic that is steadily capturing the attention of the general public, as well as those in the educational, legal and political professions. That topic is state-mandated minimum competency testing, which incorporates state-mandated student assessment programs. According to a recent summary provided by the Education Commission of the States, forty of fifty states are now implementing some form of state-mandated competency testing. Most of these tests measure academic achievement in areas such as reading, mathematics, language arts, science, and writing. Much of the controversy, however, surrounding these testing programs has stemmed from how the results are, or will be, used. Typically, results generated from such testing have been used for the purpose of determining whether or not to award high school diplomas, promote or retain, or possibly to allow for early exiting from basic programs. In contrast, assessment programs are developed to determine the need for remediation and/or reinforcement without regard to any predetermined criteria, such as pass or fail. For example, the state of Oklahoma just recently enacted a

state-mandated student assessment program to involve testing of students in grades three, seven, and ten. Their results will be used to determine remediation and reinforcement of student achievement at the district level. It is the purpose of this teleconference to address issues surrounding and implications of state-mandated minimum competency programs.

Our program will be divided into three parts. The last part will be a question and answer period during which those of you at the receiving sites are invited to call in your questions to the Telecommunications Center here in Stillwater. If you wish, you may call collect. The number to call is (405)624-7101. We have our lines staffed with operators who will take your questions and have them read to Dr. Bell by me. The second part of the program will provide our studio panel an opportunity to ask questions of our speaker. We are pleased to have on our panel three people representing three facets of education. Dr. Larry Zenke is presently superintendent of the Tulsa Oklahoma Public School System. Prior to Dr. Zenke's appointment to this position in 1976, he served as Director of Elementary Education in the Devol County, Jacksonville, Florida Schools, and later assumed the position of Deputy Superintendent of Instruction in Orange County, Florida. Mrs. Nancy O'Donnell is presently a third-grade teacher in the Perry, Oklahoma School District where she also serves as an Economic Education Coordinator. Mrs. O'Donnell was named the 1983 Oklahoma Teacher of the Year and was first runner up for the National title of Teacher of the Year the same year. Mrs. Karen Leveridge is president of the Oklahoma Coalition for Public Education and past president of the Oklahoma Parent Teacher Association. We are most honored to have our featured speaker-lecturer Dr. Terrell H. Bell. Dr. Bell is a Professor of Educational Administration at the University of Utah in Salt Lake City. As United States Secretary of Education in the Reagan cabinet from 1981-84, Dr. Bell was the highest ranking educator in the United States. This is a distinction he held once before when he served as United States Commissioner of Education from 1974-76. Dr. Bell has held positions of educational importance on both the state and national levels for over 22 consecutive years, beginning as a science teacher and athletic coach in a rural high school in Idaho. Dr. Bell is best known for his leadership of the National Commission on Excellence in Education. Twelve million copies of the Commission's report, "A Nation at Risk," have been printed, reprinted and widely disseminated since it was released two years ago. We are most pleased to have Dr. Ted Bell join us today.

Keynote Speaker

Dr. Bell:

Thank you very much, Dr. Perry. It is a pleasure to be on this program and a real opportunity for me to express some views and some ideas about state-mandated assessment programs. Before I do that, I'd like to emphasize that at the present time, nationwide, we have a great renaissance of American education taking place. There are educational reform movements all over the country. After the report entitled "A Nation at Risk" was released in April of 1983, many of the nation's governors and other leaders started to move out and campaign and press for certain reforms in education. Recommendations were made to legislatures. The National Commission report struck a responsive chord all across the nation. The report was headline news in most of the newspapers of the nation. We were

on network television as a featured story for a number of consecutive weeks and we, as a result of all of this, have had an increased attention, an increased concern, and an increased apprehension about education. As a result of this and recommendations that governors have made as well as legislative initiatives that some of our leading lawmakers have taken, we have had a number of reforms enacted by our legislatures. Some of them are quite controversial and these are pressing all of us in education to be effective in implementing the results. Along with many other forms that have been implemented, we have state-mandated assessment programs. Now, of course, these didn't come just with the national report, "A Nation at Risk." Many of them were in progress and many actions had been taken prior to that. But the enactment of legislation that affects entire states has been accelerated since the school reform movement started two and one-half years ago. The Education Commission of the States received a contract from the former U.S. Office of Education in the old HEW Department to conduct a national assessment of educational progress. This has been going on for a number of years and a couple of years ago the contract was bid again. The Educational Testing Service in Princeton, New Jersey, won the contract. They now have the mandate to carry out the national assessment program. This national assessment program is voluntary, and states, school districts, and individual schools can participate in it at their own volition. Thus, it isn't like some of the mandated state assessment and testing programs that we have. There has been a great deal of interest in these reports. Additionally, as the National Commission report came out and as we highlighted a decline in the college entrance examination scores, there was increased attention again paid to tests and measurements and how we measure effectively the outcomes of education as well as how we are able to effectively, because of these capabilities, carry out some of these mandates. So the point I'm trying to make is that the assessment effort and mandated competency testing has been increasing in intensity and has been spreading to more and more states as Dr. Perry indicated in her very fine introduction. It is becoming a major national issue.

I'd like to highlight what I've heard that's positive and negative about mandated testing, mandated competency testing, and mandated assessment for academic accomplishment and proficiency all across the country. The most common criticism, to talk about some of the criticisms first, might be of interest to our viewers. The most common criticism is, of course, that tests can't measure all of the outcomes that are desired for all of our students and so there are some serious limitations to tests and how effective they are. We all know that there are many outcomes of education. There are many desirable attainments that we want students to gain that you can't measure with an examination. We need to constantly remember that there are limitations. I will touch on others in just a moment, but at least let me emphasize that the first complaint, the one that I've heard the most often, is that testing doesn't adequately measure everything that we want to measure and that there are limitations. Sometimes we imply that a test measures everything a student has learned or attained and that this does a disservice to our students.

A second criticism is that teachers will tend to teach the test. They know that there is going to be a test given, a national assessment, a standardized norm-referenced test like the one that is required by Oklahoma law. Many others in other states also require norm-referenced tests.

Rather than teaching the subject matter, teachers, being apprehensive about how they'll be measured and how their effectiveness as teachers will be viewed, start teaching the test rather than teaching a broader and richer course of instruction. So this is another criticism that we get from the opponents of mandated testing and assessment programs.

A third comment that we hear quite often is that testing is unfair. On any given day a bright student who might not feel very well or who isn't attentive to the situation may do poorly on an examination. This might happen because of illness or emotional upset or some other condition that affects the student and so the feeling is that the administration of tests is unfair. Incidentally, I sensed some of the feeling for testing during the past summer when I visited Japan at the invitation of their national commission. They have a reform commission over there that is studying Japanese schools, and since I had been involved in the national assessment program here in the United States, the Japanese invited me to visit their program. One of the first things that I became aware of is the pressure of nationwide testing, Ministry of Education mandated-testing, in Japan. On the same day, all over Japan, youngsters will take the Ministry of Education examinations. Many students are admitted to college and their academic future, and in some ways their fate, is determined by how well they do on those state-mandated examinations. I learned there was a great deal of tension, concern, and apprehension about tests. In fact, while I was in Japan I noted a great amount of pressure and a large amount of publicity in the newspapers criticizing the testing program that the Japanese have. I think that a testing program that is as intense as theirs is would be one that certainly wouldn't be appropriate for us. Many parents to whom I spoke at the National Conference of the Japanese Junior Chamber of Commerce while I was there, young adult parents, complained about the pressure and the unfairness of these examinations. So my third point in talking about the negative side of testing is that testing is unfair. I think this is largely determined by how we use tests. If we use them for positive purposes, that is one thing; if we use the results in a negative way, that is another. Maybe we will get back to that a little bit more in our discussion.

The fourth point is one I've touched on already a little bit, and that is that testing exerts pressure and pressure will cause cheating in our schools. As youngsters respond to the pressure of these exams, some of them realize they are not going to do well and that some things are going to happen to them. They become apprehensive about it and concerned about what their parents will think if they don't do well on the examination and they start looking for a way out. Many argue that testing that exerts alot of pressure on young people results in cheating and that this is an unfortunate circumstance and another negative point that we need to consider as we consider state-mandated testing.

The other point that is emphasized, and I don't know how valid it is, is that testing rewards guessing and ability and skill to guess, especially if the exams are true/false or multiple-choice. It, therefore, depends upon chance how effective they are. Of course, we all know as we study educational tests and measurements that there are ways of handling these problems related to guessing as we handle the data on testing. But another criticism, whether you believe it is valid or responsible or not, is that objective true/false or multiple-choice exams cause this to happen.

Another criticism is that tests and measurements and the kinds of tests that we are talking about here, norm-referenced tests, measure recall and memory and test-taking skill. Norm-referenced tests therefore have a serious limitation to them. What we ought to be measuring more is the ability to reason and to solve problems, and a capacity for creativity, among other things. I'm sure many in the audience have heard these criticisms, but testing that measures recall and in effect regurgitation of facts is a common criticism of testing.

There are two more points I want to make on the negative aspects of testing and state-mandated assessment that we hear quite often. One is that tests can be used to punish or to take away privileges or rewards. When a student faces a test and we say, "Well, if you don't score to a certain point then thus and so cannot happen," this causes an attitude and a response that some educators claim has a lot of negative aspect to it. Lastly, the feeling is that tests can cause feelings of inadequacy and low self-esteem. If young people fail to do well on tests, it will hurt their self-image and will have a negative effect upon them. The panelists and other viewers may be able to add other negative points but these are the most common criticisms that I've heard on state-mandated assessment and testing programs.

Let's turn from that and talk about the positive aspects of state-mandated assessment programs. First of all, it is obvious that statewide tests give a base for comparison of how we're doing in education. If you want to know how your seventh graders, for example in Oklahoma, are doing in comparison with the neighboring school district seventh graders or with a list of other Oklahoma school districts you think are comparable in make-up, racial, ethnic, and income composition or in other states, then you will be able to take a look at how your seventh graders did and compare them with others. If you use a national norm-referenced test, as required by Oklahoma law as well as laws in other states, then you can also compare with the national norms. Giving a basis for comparison is one of the positive aspects of the assessment program. It is emphasized by many that when we measure performance, performance improves. And when we measure performance and the results are fed back to the learners with understanding, it is claimed that the rate of improvement accelerates. Thus, advocates of national assessment programs indicate that assessment will tend to put a comparative spotlight on achievement which will tend to motivate and stimulate students and educators and will cause education to improve. So the point that when performance is measured, performance improves is emphasized by the advocates of mandated assessment.

It is also argued that tests motivate students; that they like to compete and that many students like a wholesome, positive competitive atmosphere. If tests are used properly and administered properly, they can be used as a powerful motivational tool. All of us know that the most critical problem, the constant problem, the age-old problem, in education is motivation. Students will learn if you can get them to want to learn; if they have a desire and a reaching, seeking, attitude toward learning. Many argue that tests will motivate students to do their utmost. That it is challenging causes the learning environment to be more challenging and more stimulating and if you have a statewide testing program, this will be the result. I've talked to some educators in New York State who can remember the old Board of Regents exams that were administered in that

state. They are lamenting the fact that the Regents exams were taken away. They said that they had more motivation, more goals to shoot for, and more highlight events to reach for and prepare for, in preparing for these Regents exams. Those who favor statewide testing and assessment look at situations like this and say that back in earlier days when we had these that we did have better achievement and we were more motivated in that regard.

Another point close to this is that students tend to study, books come home, and preparation is underway when they are facing an examination. Of course, that can apply largely also to teacher-made exams, but the fact that students have to get ready and prepare creates a positive pressure and generates, at least the advocates claim, a certain amount of creative tension that presents positive results that are beneficial to learners.

The other point, of course, of which we are all aware is that testing and standardized testing can tell us how to improve our teaching if we analyze it in considerable detail. As we see others who do exceptionally well in the testing program, we can seek out those schools and learn what they are doing that works so well. When we can find our faults and our failings through testing, it is argued that this helps us and can be an effective part of the school improvement program. Advocates believe that this will be the result of the newly mandated student norm-referenced assessment program in Oklahoma. Many will be searching the details of the students' responses and out of that will come some improved teaching which will benefit students largely.

Another point is that tests can be used to teach the results of tests, the effects of tests; that you can have effective follow-up after a test is given. After a statewide exam is given you can use some of the test items as a source of discussion and you can ask students what were some of the tough items; what were some that they didn't know. Out of these can come teaching opportunities. It is also emphasized that tests tell us how much progress we have made and if tests are given regularly we can compare them with previous accomplishment that has been attained. Down through the years in Oklahoma, we will be able to look back and see what has happened. Of course, we have had that comparison nationwide with the college entrance examinations. The ACT and the SAT tests are given all over the country and are largely the two major exams that are given to students who are going on to college. We are able to compare scores for nineteen or twenty years to determine how students are doing today compared to previous years. The arguments are that, in effect, we'll be drawing a line; we'll have a benchmark from these exams that are given this year in Oklahoma and that that will provide an opportunity to measure and to know how much progress we have been making in that regard.

It is also argued, talking on the positive side of state-mandated assessment, that test results can generate support from parents and from the public if they are used properly to identify the needs that we have. If used in conversations with parents, if used in reporting to those who are patrons of support of school, we can rally support. Many have said that this was the best thing that has happened and many have said it was the worst thing that has happened, so I don't know how to assess it. Talking now on the positive points, many have said that one of the best things that has happened is when we released the state-by-state comparison of how

graduating seniors did in 1983 and 1984 on the college entrance examinations. Some of you will recall that while I was Secretary of Education, we released these data in response to requests from a large number of governors and we ranked the states and how they did in college entrance scores. Many have said that this generated support. Some have said they were able to use that in going to their legislatures to get more support for their states. Some argue that that will come about in schools and that actually if your school system or if your individual school does not do too well on an exam that you can use that for persuasion with the parents to get more help, more attention, more financial support, and more resources for the schools. Some say the opposite of that can happen. If we do poorly, we'll be criticized. Parents and the public will put us down and will want to withdraw their support and attention. So that argument cuts both ways, but at least those who are favorable to statewide mandated assessment programs argue that we ought to be using test results in that regard.

It is also argued that test results can be used to generate friendly competition from one school district to another. I'm reminded of that use of friendly competition as I visited the Devol County schools in Jackson, Florida, a couple of years ago. They had an academic super bowl down there. In that school system they are constantly quite openly using friendly, hopefully positive, approaches to competition as a means of getting students to pay more attention to their academics and to strive and try to apply themselves more effectively. Without publishing the results on individual students where the situation would be embarrassing, you can publish the results schoolwide and promote wholesome friendly competition among schools, at least the argument is that we can do that. This would function much academically like the athletic scoreboard does. If you have ever been to a Little League ball game in an area where they don't have a scoreboard, you have to walk around to find out what the score is. Many say that if we took all the scoreboards down, attendance at athletic contests would be gone. Those who argue for statewide assessment argue that it is tantamount to the scoreboard; that it will generate interest; that it will generate friendly competition that will be helpful and will cause students, teachers, parents, and others to pitch in and try harder in education.

Another argument on behalf of statewide mandated assessment is that we need to measure performance and we need to report performance because that is life. We live in a free enterprise competitive environment. We live in a world that is very competitive and by learning how to compete in school, how to perform well academically, we can then prepare ourselves to face our future and to face the reality that is going to be out in the real world. We all know that as we begin to compare ourselves that there are greater and lesser persons and as we think about this we need to acknowledge that and be able to adjust ourselves to it. Whether you accept that as an argument or not, and I'm trying to present here in my opening presentation all the arguments that I have heard, one of them is that living in a more competitive environment in school prepares you for life. Critics say that we have sheltered students from this. This is the other side of the argument that I gave earlier when I was talking about the negative side of testing where it damages self-esteem and hurts the will of many to want to try to do their best.

Another argument on the positive side of test results is that they report progress which we can use in our efforts to gain more financial support for schools. We can take the results and if we highlight them and compare them with those of other schools which are providing more resources and therefore attain higher achievement, we are going to make the progress that we need in getting support for education. Many argue that we ought to use test results; we ought to openly report them; we ought to publish them in the newspapers school by school and school district by school district; and then use them to generate more support for education. Whether that is valid or not, I'm sure many of the viewers have heard points on both the positive and negative side of this question. You could say that this is a valid point and that isn't one or that point is totally off base, but I thought by highlighting some of these prime arguments about state-mandated assessment, that we would have an opportunity to examine the questions a little bit more.

Let me turn from that now that I've touched on the positive and negative points of state-mandated testing and express some more general views. The first thing that I would emphasize, and we are well aware of it, is that parents, taxpayers, and the general public are expecting us to do better in education. I'm not one who accepts the proposition that education has done poorly. If we compare our resources, if we compare some of the handicaps, difficulties, and limitations that have been placed upon us, heavy class loads and so on I think that American education has performed very credibly. I'm not one who believes that we ought to be moving forward with some kinds of mandated efforts such as mandated-assessment as a way of fixing some terrible, scandalous deficiency in education. I do believe, this is my own view, that the positive items outweigh the negative as far as statewide mandated assessment is concerned. Many, but not all, of the negative points can be avoided if intelligent people work at it. But I would emphasize that I think that if poorly explained and foolishly used, test results that come out of this assessment effort will do more harm than good.

I'm convinced that the people of Oklahoma and other states have the capacity, the sophistication, and the capability to handle a mandate for statewide testing and assessment such as the one that is now being implemented in the great state of Oklahoma. The public needs to know and I think the more we tell the public about how well we do the better off we are. I think we will have more precise and comparative information if we are able to carry this out. If schools aren't doing too well, they are the ones, as I said before, that will have the most to gain, I believe, if the results are explained in an intelligent, rational, dispassionate way. We need to realize that if people are not constantly informed about what tests can do and what they can't do, if we don't do a good job of education about tests and measurements and assessment, then the results can be distorted and the harm can happen. Many of the negative points that I emphasized in the beginning of this presentation can be the result of this kind of a statewide assessment effort. But intelligent people, if they have effective communication with parents and others, can handle the results. Ignorance has always been the age-old enemy of education and ignorance in testing and ignorance in communicating results, indulging in what I think is the favorite American pastime, jumping to conclusions, could cause a state-mandated assessment program to do more harm than good. But I don't think that is going to happen. I'm confident, for example, that

the educators in Oklahoma and in other states where they have these statewide assessment programs will be able to handle the results intelligently. They will be able to communicate the results. They will be able, in a nondefensive open way, to share with parents what the problems are and get parents involved.

We all need to keep in mind that it isn't just the school that influences achievement. Sometimes some of us are a bit too defensive about how our schools are doing. We are a bit too reluctant to tell the public that we have problems and we have deficiencies and that we have some areas that we are not getting across what we ought to be getting across to our students. They are not attaining the mastery of the academics that we would like them to have. In that situation, I believe that we hurt ourselves if we hold back the results. I'm one who believes, based upon my experience, that the more we inform the public and the more we tell the parents, the better off we are going to be. The point I wanted to emphasize is that the home is the first classroom of every child and the parents are the most powerful and influential teachers. We need to get the homes, the parents, and the communities behind our schools and if we use testing intelligently, if we apply the results, if we interpret the results, and if we are broadminded and mature and open about it, I think there can be more good than bad come out of a statewide assessment program.

I would emphasize that there is going to be a lot of apprehension, and there is, when a state moves into these kinds of programs. You can't help but worry about who's going to interpret what and about how someone is going to, in some almost malicious way, bring about an interpretation that you didn't intend and that you didn't want. But by and large, I'm convinced that statewide assessment programs, let me repeat again, if intelligently applied can be used to advantage. I think school superintendents, principals and teachers can turn assessment programs into opportunities for them to generate the kind of support that we need for education. I have a lot of concern and share in some of the apprehensions that there are. I have been one who has favored more statewide standardized assessment of education and the utilization of the results to generate support, to diagnose learning difficulties, and to bring about efforts that will cause achievement to improve. I'm convinced that all of these outcomes can become a part of an intelligently administered statewide assessment program.

Let me emphasize in closing that I think that the opposite can happen. I've seen the results used in less than an intelligent, positive way and we can let students down. We can damage their self-esteem if we are not intelligent and professional in how we utilize the results. Largely, I think it is going to depend, in Oklahoma and other states, on how effective we are in implementing the laws, how intelligent and broadminded we are in interpreting the progress and carrying out the mandate. That, then, is a summary of my presentation and I understand that we now have, as I turn the time back to Stillwater, a question and answer session lined up to follow. Thank you very much and I'll look forward to responding to questions.

Dr. Perry:

Thank you, Dr. Bell, for sharing some of the complexities of the issues that surround competency testing. We now are getting ready to enter our second phase in which we will have a question and answer period with the members of our studio panel. I might add that the lines are open whereby those of you who want to call in from the receiving sites may do so. The number again is (405)624-7101. Starting with our first panelist, Dr. Zenke, would you now pose your question to Dr. Bell.

Panel Discussion with Dr. Bell

Dr. Zenke:

Thank you, Katye. Dr. Bell, I feel that you very effectively summarized the strengths and weaknesses of state-mandated competency testing across the nation. I think you also very effectively summarized your closing comments with your very strong position in support of such state-mandated competency testing programs. As I look back approximately twelve years to when the Florida legislature passed the Omnibus Education Act which included minimum competency testing for students in the state of Florida, and then as I've watched since the passage of that act in state after state across the nation, various other legislatures having also passed such legislation, and then as I look back to two and a half years ago when the National Commission on Excellence in Education released its report for which you were responsible, "A Nation at Risk," which again brought about a rash of reform state by state as mandated by legislatures across the nation and all so with the very act of involvement of governors, and then as I compare those mandated reforms brought about on a statewide basis with a couple of other approaches to reform--one is described by Dr. John Goodlad in his recently published "A Place Called School" in which Dr. Goodlad says the individual school is the largest and most appropriate unit for change, and then as I refer to another book by Carl Marlberger entitled "One School at Time," Dr. Marlberger being a former Commissioner of Education in the State of New Jersey and now involved with the National Committee for Citizens in Education, he says as also stated by Dr. Goodlad that the individual school is the largest unit for change--do you see a contradiction in these two approaches to change, one being the state-mandated approach to change and the other the individual school, school-by-school approach to change, whether it be on minimum competency testing as it addresses that single issue or the greater breadth of reform we've seen across the nation?

Dr. Bell:

I think, Superintendent, that you have made a very good point, and you have raised an excellent question. I think there is potential here for a conflict. I think that it will largely depend on how the principal and the parents and the teachers in the individual school might use the statewide results. Another thing that could happen, and I worry about this, is that those who are not knowledgeable, well meaning but not knowledgeable about education, can start asking for the results and begin to try to do other things on the state level, or even on the school district level, that are not sensitive to local needs. I've seen curriculum development on the floor of the house and the senate in legislatures all across the

country--well meaning legislators, good strong pro-education minded legislators--but some mandates just shouldn't come from the state level and I worry about that. I think the trend has to be towards more local autonomy. I even think that from the school district level, and the central office level, the principal, the neighborhood community and the faculty need more latitude, more encouragement and more opportunity to lead out. They need to do that within a framework of state law and school board regulations. My fear is that the state law and the school board regulations can get to be too heavy handed. Both Marlberger and Goodlad make splendid points in that regard. I wouldn't say that their points per se would invalidate the case for a statewide assessment program but I think a misused statewide assessment program could result in that. We would start driving, directing and mandating things from the wrong level and so I agree with the point you've made, Superintendent. I think it ought to be a caution that all of us would follow. Do you have another question, Dr. Perry?

Dr. Perry:

Yes, I do. We are going to turn now to Mrs. O'Donnell--Nancy.

Mrs. O'Donnell:

Thank you, Dr. Perry. Dr. Bell, you are certainly to be commended along with the National Commission on the focus your work has given to the critical issues in education and the resulting forms. As a classroom teacher, I am concerned about state-mandated competency testing and I share some of your concerns in this light. If some are determined competent, then are some to be determined incompetent? After students are determined incompetent, what happens next? In a time when funding at all levels is a critical issue and staffing with certified personnel is rapidly becoming a critical issue, who will redesign and administer the programs for those declared as not competent? And who's going to pay the bill? I wonder if we dare to attempt to diagnose students as educationally ill and not determine a proper remedy and not administer that remedy and not make every effort to allow that students become educationally healthy. In my time I've seen too many miracle cures go for nought because only the diagnostic and prescriptive stages have developed. No one really wanted to pay for the medicine. Do you have some ideas about what we are going to do after we have determined competency and, of course, that speaks for itself? Some will be determined competent, some incompetent. What are we going to do?

Dr. Bell:

Nancy, let me say first of all, it scared me a little bit when I found out that the state teacher of the year and national runner-up was on this panel. I just want to express my admiration to you. That's tougher than being elected governor of Oklahoma, to win that state teacher of the year. I realize that you make a good point, and that is that if we have competency testing and we start saying this one is competent and this one is competent, pretty soon we get down to where we say this one is incompetent. The point I was trying to make is that if we start using it in that way, we're going to destroy ourselves in the process. I think we have to beware of labels. I think if we can identify educational problems that we

have and then bring about solutions in a positive way, we're going to be helpful. I learned something a long time ago in my first assignment up in Idaho, our northern neighbors here in Utah. I grew up in that state and worked there in my early years and I met a young woman up there. I noted that she was top of the class, but according to the IQ tests she shouldn't have been. She applied herself very, very conscientiously. She has always been an inspiration to me. She wound up valedictorian of her graduating class although at one time in her career, I think she could have been labeled as incompetent. Another point that I've learned is that Albert Einstein, when he was a school student over in Europe, was labeled by his teachers as a nincompoop and as incompetent to learn. So we need to be careful about these labels, Nancy, and I think that is the point that you are making. If we misuse the results, that's going to happen. I think we have a great potential for harm as well as good in state-mandated assessments and if we use the results in a positive way, if those who pass the law to mandate the assessment in Oklahoma, for example, will look at the results and say that we need resources here and we need them there and we are willing to put these resources up to solve these problems, then we will have education marching along the way that it ought to be. If on the other hand, we take the results, Nancy, as you mentioned it, and apply them in some negative way by labeling incompetent, then we will do a lot of harm. I'm presuming, and maybe some people will say I'm getting naive in my old age, but I'm presuming that we've reached a point, we are intelligent enough, that Oklahoma people are intelligent enough, to handle the kind of information that is going to come out of this. If they are not, if that is an erroneous assumption, then heaven help us because we are going to have all kinds of difficulties. Now, do you have another question, Dr. Perry.

Dr. Perry:

Yes, we are now going to turn to Karen Leveridge.

Mrs. Leveridge:

Thank you. Dr. Bell, it is wonderful having you visiting with us, though by long distance, today and we welcome you to our state. Those of us that have worked a very long time to support public education in every possible way accepted the report from the Commission on Excellence directed by you with a great deal of applause and concern at the same time. It was time to bring those problems to the attention of everyone across this nation and that commission did an excellent job under your direction in seeing that we became more aware of the concerns facing public schools today. There are some pressures certainly in the world as we live in a faster world day by day and I'm a little concerned about some of those things that may put additional pressure on not just students but the teachers as well. As student assessment programs are mandated at a rapidly increasing pace across the country, there are some concerns that come to mind to many of us who work in a voluntary manner with the public schools. For example, the teacher preparation of the students to take those tests is a job that carries much responsibility and hard work with it. Then following the tests which those students take, there seems to be a common sin placed upon the teachers if there is failure of those students, or any part of the students, who are in the classroom. How do you respond to the added pressures that are put upon our classroom teachers today as they not only

prepare students for the increasing numbers of tests, but as they also face the results of those tests, and oftentimes blame for the failure, and I fear all too infrequently receive praise where success comes about?

Dr. Bell:

You make a good point, Karen. I'd like to commend you for your long dedication to education not only in Oklahoma but on the national scene with PTA. We need more people like you to help us solve our problems and its a pleasure for me to be able to say that as we are on this panel together. Pressure is a double-edged thing. We talk about having creative tension, underline that word creative. A little bit of tension, if it is creative and positive, is good for us; it keeps us up on that outer edge where we ought to be. But teachers, many many teachers, labor under horrendous loads. They have an impossible task because we give them such heavy class loads and we dump on them so many requirements. We've expanded the secondary school curriculum to the point that it is just impossible to get mastery learning as a result of it. I might just mention also that I heard (I referred in my prepared remarks to my experience when I visited Japan) a lot of parents who were uptight over pressure. As I talked to students, I talked to a high school principal and he told me how they feared that national ministry examination. I dread having that happen. Now we looked at the Japanese and they had a very impressive education system over there, very intelligent people. They are working hard. I thought there was much we could learn from the Japanese education system. But I wouldn't trade them because of the pressure cooker situation that those students labor under. Parents are up in arms about it. That is why Prime Minister Nakasony has appointed this commission. They are not talking about what we are talking about here. They are talking about taking some of these pressures off, so I know, Karen, that we can overdo that, not only on students but on teachers. We have some teachers who carry some horrendous loads and someway, somehow we need to get the resources and the support so we don't have teachers teaching so many students, so they have more time to plan and prepare and do a better job of teaching. We have been saying that all the years that I have been in education, and hopefully, out of this, out of the Commission report and reform movement, out of actions like, for example, the Oklahoma legislature to mandate this assessment, I hope we will get the information so we will get a more reasonable response to our needs, and I worry about that. I'm fearful that our efforts are going to peter out and not yield the benefits that we need. So I join you in expressing apprehension about this. These tests can be used as a pressure point and can be harmful if we are not very wise and intelligent in how we utilize them. Now, do you have any more questions, Dr. Perry?

Dr. Perry:

Yes, I have. In fact, I have several questions from some of our receiving sites and I will address some of those now. One question comes to us from East Texas State University and the question is: Are statewide tests a movement toward a national curriculum?

Dr. Bell:

I hope not. I don't think we need a national educational curriculum. I do think we ought to be comparing our results and learning from each other, nationwide, but I don't think we ought to be looking at a national curriculum. Now I think there can be a nationwide consensus over what students need to learn, basically. The National Commission on Excellence in Education in the report, "A Nation At Risk," as you recall recommended that all students who graduate from high school study four years of English; that they study mathematics for three years; that they study science for three years; and social studies for three years; and they mandated that if a student is going on to college, they ought to have two years of foreign language. Now that eighteen-member commission, a very distinguished, prestigious panel, had a tremendous number of debates over that and they took testimony all over the country. But they came to a unanimous conclusion on that, but they did not feel that we ought to go from there and mandate the textbooks and the curriculum content that we ought to have. I hope we never come to that. That is why we ought to avoid having a Ministry of Education, European style or Japanese style. It just doesn't fit our country. But education is a state responsibility, and the prime policy-making body is the state legislature. I think it is appropriate for it to come from that level and I would like to emphasize, if I can, a little parenthetical point. I have said many times that education is to state government what national defense is to the Federal government. As you see Ronald Reagan pressing for those budgets you will notice how he fights for the defense department budget. Well, I've wanted governors to fight like that for education because education is to them what national defense is to the President. That's why I've been pleased to see many governors, especially in Texas and Oklahoma recently, coming out like they are for education. We need more of that but we certainly don't need, to get back to the original question, we certainly don't need a nationally mandated curriculum. We ought to leave that at the state level. We can get together, we can meet, we can talk across state lines, we can have national and nationwide meetings, but we surely don't need a national curriculum or a federal ministry of education, in my opinion. Do we have another question, Dr. Perry?

Dr. Perry:

Yes, I have. This particular question came in from the Oakland school system in Pontiac, Michigan. Dr. Bell, would you discuss the impact of state mandated programs on the autonomous control by local school district boards of education to determine curriculum and what should be taught in their community?

Dr. Bell:

Yes, that is an excellent question. I think that if the state-mandated assessment program gets down and is put into state law to where it starts prescribing minute detail, then I think that this can begin to happen, this untoward result that was just implied by that question. From what I know about the Oklahoma Law, it specifies, I think, five or six subject areas, but they are broad subject areas that we all teach--mathematics, sciences, English, and I think citizenship is in there. Those are broad areas that are common to all schools and if we will keep our state assess-

ment requirements broad and general, then we are not going to be impinging upon the autonomy of local school boards and local communities. We will still be able to do some things in a way so we will have common data and common opportunities for comparison. I wouldn't want to say that this isn't something to fear. Anytime, anything can happen when the legislature is in session. The former secretary of HEW, Elliot Richardson, used to say no man's life, liberty or property is safe when the legislature is in session, a kidding comment. But the point is, when the lawmaking bodies are deliberating, they need to be very prudent and careful and wise in what they do. They need to be careful about specifying detail in such a minute way. So that is my feeling about it. I think there is a potential for this problem happening but I don't think it is going to happen in our states because I don't think the educators will let it happen. I think the powers of persuasion will be such that that won't take place, but it was surely good to hear from Pontiac, Michigan.

Dr. Perry:

I have another question here, in fact it is from a librarian or a media specialist from Forgan Public Schools. This person's question is: I am reminded again of the IQ testing controversy of several years ago. Is there a national normed test that will truly test the knowledge of a student who has spent an entire life in a small rural community removed from any major metropolitan area by more than 250 miles? Are we going to once again mistakenly identify another group as deprived or ignorant as a result of cultural differences?

Dr. Bell:

I hope that we won't and I don't think that we will. I think students in this isolated community that we talked about, especially with modern technology like this that we are using and I commend Oklahoma State for the use of this advanced technological capability, need to measure and compare their mastery with others. We can find comparable schools. There isn't only one school that is in a rural isolated area, there are thousands of them all over the nation. There wouldn't be only one in Oklahoma or one in Michigan or one in Texas. If we can find comparable schools, comparable institutions, we can get an understanding of how well we are doing as related to others with similar circumstances. Additionally, I believe that we can expect students in rural, isolated areas, who may not be quite as culturally privileged as if they were living in cities where they would have symphony and opera and all of those things, still need to be inspired and motivated to get a good basic education and as long as competency testing measures that, I don't think that we are going to get the unwholesome results that we fear. There is always the danger of it and there is always the possibility of this happening, but there are advantages in living in a rural community. I grew up in the unlikely sounding place of Lava Hot Springs, Idaho, population 700 and I had some advantages as well as disadvantages in attending that school. As I reflect upon it, I was pleased that I went to a small school. I think that the potential for mastering basic subject matter, getting a good command of mathematics and English and social studies, can be attained even in a rural and isolated school setting. Hopefully we will be successful in those kinds of endeavors. Do you have another one now, Dr. Perry, you would like to raise?

Dr. Perry:

Yes, I have one more from a receiving site then I will readdress our studio panel here. The question, again, from Pontiac, Michigan is: How do we deal with the fact that some students even though capable do not test as well as other equally bright youngsters?

Dr. Bell:

Well, this is a reality. Some students just don't do well on tests. Some freeze up when an examination is given and I think in those instances we need to be looking for exceptions and we need to be learning how to manage the exceptional situation. I think alert, well-prepared professionals can find, detect, and remediate these kinds of differences. If we just go ahead and dogmatically administer tests and we see a test result and we know that it is absolutely not representative of this student's achievement as we know it, this is where we need to be looking at results and scrutinizing them. Not only do we need to be testing students but we ought to be questioning and testing the tests. And this needs to be constantly recognized and reconciled as we consider the differences. I don't know if I still responded to that Pontiac question effectively. It is a very insightful one and one that we ought to be thinking about more. The whole thing about testing is that it can be state-mandated testing but it is a broad-brushed approach and we need to realize that. We are still dealing with individuals and we need respond back to that. Well, thank you, Katye. It has been a pleasure to have that interaction.

Dr. Perry:

Oh, sure. We are going to turn back now to our studio panel. Dr. Zenke, do you have another question that you would like to ask Dr. Bell?

Dr. Zenke:

Dr. Perry, before I ask the question I would like to throw in a commercial relating to the second to last question that you received. On all the other questions you have identified the city and the state and that particular question you simply said it was from Forgan. I'd like to explain to Dr. Bell that Forgan is a small community in northwest Oklahoma. I'll be eating steak at one of the finest steak places the Old Railroad Depot there the first part of December as I travel out there to go pheasant hunting so I invite you to have a steak dinner with me and go pheasant hunting.

Dr. Bell:

Oh, that would be great. Good to see you, Superintendent, in Tulsa, incidentally.

Dr. Zenke:

It is good to have the chance to interact with you again. The question I would have for you would be in the area of norm-referenced as contrasted with criterion-referenced tests. Most educators and parents across the nation understand norm-referenced testing. We have been administering na-

tionally developed norm-referenced achievement tests for the past many years. Norm-referenced tests by definition compare one student against the other with, again by definition, 50% in the normal population being below the average and 50% above the average and none of us want our students to be below the average. Criterion-referenced tests, again by definition, compare the individual student against a set of predetermined objectives. What is the place for both types of tests within a school system being administered or as you would view it, is one preferable over the other?

Dr. Bell:

I would like to see us use more criterion-referenced tests. Now, by responding that way, Superintendent, I don't suggest that we abandon the norm-referenced tests. I think we need both but we've spent a lot of time on the one and not enough on the other. I think we ought to do more to help our teachers become more capable in teacher-made tests and many of them, of course, are criterion-referenced. There is an art, a sophistication, to developing good tests. I think we need to do more in teaching educational tests and measurements on our campuses so that teachers can learn how to develop good tests, how to interpret the results, and how to go through and identify the items that don't measure want they want. I just say that we have a long way to go in criterion-referenced tests and I think we should spend more of our resources on them.

Mrs. Leveridge:

We've not talked a great deal, Dr. Bell, about funding today and there are some interesting questions that come to mind as the result of recent research that tends to dispute the theory that expenditures have no effect on student outcomes. One report that comes to mind that I've seen was in the Harvard Educational Review of November of 1984 by Powell and Steelman. I just wondered as we talked about this many faceted issue if that research is kept in mind. Are we faced with a continuing need to increase per student funding in order to raise our test scores and have them where we would like them to be?

Dr. Bell:

I think what does relate well to student achievement is teachers and how effective they are. Consider the bright, talented, and capable young people that we don't get into teaching because teaching isn't competitive--we don't compete in the human talent marketplace--and a lot of our able young people who ought to be in teaching are going into engineering, business, law, and medicine. That is where I think the funding comes in. Another place where I think we begin to lose, regardless of the research that tries to establish a relationship between funding and achievement, is when our able teachers, many of them, leave the profession. A study done in North Carolina was a shocker to me a couple of years ago when it indicated that the best and talented teachers, significant numbers of them, were leaving within the first five years. We have a terrible attrition rate there. I think that is what is important as far as funding is concerned. Most of the schools all over the nation are starved for funding. We surely are in Utah. We're fiftieth now in the nation in expenditures per child in education out here where we take a lot of pride in education. I

wish I were talking on a local Utah network right now. You have it in Oklahoma. You are up there a ways but you are not where you ought to be in funding. No one can tell me that funding isn't related to quality in education. It simply is. It is related to the quality of the teacher; it is related to the kind of human talent that you get into teaching. If you look at some of these obsolete books in some places where I've been around the nation you realize that we're hurting and starving for funding. The teacher load is too heavy. The classload is terrible nationwide and that could be solved by funding. So, no matter what they want to say about it, the opponents to it, we have a funding problem all across this nation. Let me just say, and I'm probably talking too long on this, but in Japan they have set teachers salaries at 30% above the national civil service level. If you teach over there, you are in the upper 25% of the economic status. Over here, we are 30% below the federal career civil service level. Here is a priority we've set that relates to funding. Over there it is a national policy that teachers are going to be funded on that level, that their salaries are going to be at that level, and I noted how many bright and talented young people were struggling and competing to get into teaching. Over here we're sort of in a beggar's situation to get them. I emphasize as impassioned as I can that we need funding and don't let anyone tell you that that isn't important. I hope that this national drive, and I hope that this statewide assessment program you now have in Oklahoma and we have in other states, will give us the entree and the means of showing that funding is significant because it is. They say that money isn't everything, but it surely is important in all parts of our society and let me tell you it is as far as American education is concerned. We are either going to invest in our schools or we are going to be disappointed and are not going to maintain our proud future in this country. You can tell I'm getting on my soapbox on that one, Karen. You triggered a sensitive point with me. But I just can't believe anyone who says that we don't need funding. The states have the responsibility. We have too much load on property tax and we need more coming out of a good balanced state funding program out of our state legislatures.

Dr. Perry:

Dr. Bell, we have one final question here. It comes from Raleigh, North Carolina, from a professor of education there. That person's question is: Using test results for graduation or grade promotion is labeling. Are you saying this is a misuse of state-mandated testing?

Dr. Bell:

I don't know that we need to use to use test results for that purpose, although, I think that if we do find students--this may be in conflict with what I just said--who aren't prepared to do work on the next level, if there are ways to help them to solve those problems, that is where testing can come in. If we can identify those specific items in which a student lacks competency, we can utilize that to help them move forward. Again, if we are using tests as punishment or for downgrading, we are using them for the wrong purposes. That is what I've tried to emphasize in this presentation. I don't know how effective I've been at it. We need to be positive and upbeat in the utilization of test results. There is a place for us to identify a threshold level of competency but then, in a positive way, to help an individual reach the level where he/she ought to be; not

to dump shame and failure over him/her. So, in all instances, we need to use good pedagogy and sound motivational practices in all that we do. That is my response to that.

If this is the conclusion, I would like to say, Dr. Perry, members of the panel and all of the audience, that it has been a pleasure to be on this interesting show and a privilege to participate with you. I'll be leaving Salt Lake City now to run to get on a plane. I'm going to join the University of Georgia tomorrow in the celebration of their 200th anniversary in Athens, Georgia. Thanks so much to Oklahoma State for inviting me to participate in this very interesting program.

Dr. Perry:

Thank you, Dr. Bell and do drive carefully. I hope that today's program has helped you to better understand the issues that surround minimum competency testing. We are very pleased with the responses that we've gotten from you and our receiving sites and of course from our panelists here. As would be expected, time would not allow us to examine all of the issues that surround this particular topic. For instance, there may be some philosophical issues that need to be addressed as well as some more indepth psychometric kinds of issues that need to be brought up as relates to minimum competency testing. We are planning an all-day conference here on campus tomorrow to address more specifically the educational, political, financial, and psychometric, as well as the philosophical issues, that surround minimum competency testing. We will have, as an example, Dr. Thomas Fisher who is from the state of Florida and, more specifically, is the Director of the state assessment program there, who will describe to us Florida's experience in implementing competency testing. We will have Dr. Beverly Anderson, who is from the Education Commission of the States in Denver, who will address the psychometric implications surrounding competency testing, and then we will have several faculty members who will be in the position to describe more indepth these basic four issues that surround competency testing. As Dr. Bell was addressing the topic, I think it is clear that what he is supporting is more of an assessment-oriented program as opposed to a minimum competency testing program in terms of what uses are made of the results of those testing programs; and that is to use them more or less for the purposes of remediation and reinforcement of the issues surrounding minimum competency testing as opposed to using them to determine the awarding of a high school diploma, to determine if a person should be promoted or retained, or possibly should be granted early exit from the basic courses. I am quite pleased with the support that he has shown for the area of education in terms of the confidence that he has in our whole institution or our whole system. With that, I would like to again turn to our panelists, Dr. Zenke from the Tulsa Public School System, I say thank you. To Mrs. O'Donnell from Perry, and then to Mrs. Karen Leveridge from Oklahoma City, thank you all for the support you have given and the time taken away from your schedules to join us in discussing this issue. To those of you at our receiving sites, thanks so much for your support and I hope we have addressed some of the more important questions that surround the issues of minimum competency testing.

Using the Psychometric Aspect of Statewide Student Testing to Promote School Improvement

by

Beverly L. Anderson

Public education is in great danger. Over the past decade schools have increasingly lost respect. However, the reform movement of the past two or three years has rescued us briefly, giving us a chance to improve schools. The public is ready to desert the schools if we do not capture this moment.

In 1985, our buzzword is "window of opportunity." William Shakespeare said it a little differently, hundreds of years ago:

> There is a tide in the affairs of men
> Which, taken at the flood, leads on to fortune; ...
> On such a full sea are we now afloat;
> And we must take the current when it serves
> Or lose our ventures.

How can we recapture the public trust and move ahead toward excellence in education? Actually, there are several rays of hope. One of the brightest is emerging from the research on effective schools, schools in search of excellence and equity. I would like to use some of the findings of that research as the context for our discussion today of the psychometric aspects of statewide testing of student competency. After a brief description of that context, I will explain how choices made in the design and use of a statewide testing program can actually determine whether such testing promotes or detracts from our quest for excellent schools.

During the 1970s, a number of research studies identified characteristics of effective schools. The studies varied widely--some were outlier studies that separated the best schools from the worst schools on statistical grounds and then tried to account for the differences. A great many were case studies of schools that seemed effective in terms of student outcomes. Others evaluated particular programs--reading programs, for example, or compensatory education programs--in an effort to find out why some programs were especially successful. There is now a strong research base that describes the characteristics of effective schools and classrooms.

I would like to call your attention to four characteristics of effective schools that are particularly relevant to state testing.

1. Expectations are high for all students. Teachers, principals, and everyone else with a part to play in the education process expect all students to be able to learn basic academic skills.

2. Instruction is guided by a defined curriculum that is clear and focused. Learning goals and objectives meet school and district guide-

lines and are approved by teachers. Principals play a key role in ensuring a strong focus on learning.

3. Student progress is closely monitored. Teachers keep careful track of how well students are doing. So do schools and districts.

4. An atmosphere of collegiality and improvement exists. Collegiality among teachers and administrators and involvements of parents and the community promote a positive environment for students and educators alike. An atmosphere of striving to improve instruction and learning is present.

How, then, can testing promote such characteristics of school effectiveness? We have all heard the many arguments about the negative effects of testing. Here are some:

1. Testing promotes rigidity and lack of individual attention to student needs.

2. Testing promotes unhealthy competition among teachers, schools or districts rather than promoting collegiality.

3. Testing is based on and builds distrust. Administrators, policy makers and the public do not trust what teachers are doing in the classroom and thus require tests to be sure teachers are doing their jobs.

4. Testing creates a climate of fear and a bureaucratic mind-set.

Yes, testing can have these effects but we can choose to structure and use testing in ways that avoid these effects. Indeed, we can choose not only to avoid negative effects, but to promote a positive situation where testing actually plays a role in promoting effective schools.

Thus, today, I want to focus on testing--state testing--that (1) promotes collegiality, (2) encourages high expectations for students, (3) builds schools that are exciting places to work and learn, and (4) keeps the school clearly focused on the goal of academic achievement. I want to look at how testing can be used at the cutting edge rather than the trailing edge of the curriculum. The key to moving testing into these roles is the way the test planning and development are done--the process of handling psychometric and technical aspects of testing. I will particularly focus on contrasting practices commonly used today with practices that would more effectively promote such a situation. Although there are many ways we could approach this matter, I wish particulary to discuss the differences that would exist if we look at state testing as a tool for curriculum advancement rather than accountability.

Differences in Test Development Depending on Testing Purpose

All too often testing programs are begun with insufficient thought to the program's purpose. The purpose is often described as to help improve education and/or to help students and educators know what skills need to be taught. Such broad descriptions provide little guidance for test design. Several ways of defining or distinguishing testing purposes exist that do

help guide test-development decisions. For example, tests can be distinguished on the basis of whether they are intended to:

1. help a teacher provide appropriate instruction for a student;

2. determine if a student meets a predetermined criterion for selection into a special program of credentialing; e.g., high school graduation; or

3. evaluate an education program.

Another way of distinguishing tests, especially state tests, is in terms of whether they are to be used to:

1. monitor changes in performance of certain groups of students over time to assist in program planning;

2. hold students, teachers, and/or administrators accountable for certain levels of performance; or

3. help curriculum specialists, teachers, and/or administrators determine how to improve the curriculum in areas where skills are not currently well taught either because the skills have not been understood well enough to allow effective teaching or because the teachers have not been trained in teaching those skills.

Many statewide assessments began as monitoring devices simply to determine "what is" in terms of student learning. Once results were reported, people quickly leaped into a judgmental frame of mind--"the achievement levels are too low. It must be the fault of...." Different groups would place the blame on different parties but the most likely ones were the teachers or the students. This mentality quickly shifted us into an accountability approach. And, unfortunately, people were frequently being held accountable for skills they had not been told they were responsible to learn (students) or responsible to teach (teachers). This approach has tended to move us away from at least two very important characteristics of effective schools:

1. high expectations for all students; and

2. an atmosphere of collegiality and improvement.

We ended up moving toward minimum competencies--which did raise expectations for the lower performers but unfortunately had no effect or, worse yet, the opposite effect for high achievers. We also began to create an atmosphere of distrust and rigidity in many schools.

With just a few changes in our approach to testing we can turn things around and move toward tests which advance the curriculum and create a more positive climate in the schools.

By looking at differences in the psychometric characteristics of commonly used state tests that emphasize routine monitoring or accountability and ones which are oriented more toward curriculum advancement and improved school climate, we can better understand how important it is that we

define clearly our purpose for and orientation toward testing. We will then be in a much stronger position to design tests that best accomplish their intended purpose and, overall, promote effective schools.

Before getting into some of the technical details of the test development process, I would like to give you two examples that illustrate fairly desperate points on the spectrum of state testing--one that is heavily accountability-oriented and one that is oriented toward curriculum advancement. Neither represents extreme ends of the spectrum but rather illustrate the considerable distance that exists between today's common practice with its accountability emphasis and a realistic one that would emphasize advancing the curriculum and promoting a collegial atmosphere.

The typical state assessment test today covers reading, mathematics and language arts. Basic skills are particularly emphasized. It is developed by the state education agency involving an advisory group composed of curriculum specialists and teachers. Multiple choice items are developed that measure what is commonly taught in most classrooms at the grade level to be assessed.

All students at the grade level are assessed in the spring. Scores are reported separately by skills. A criterion such as three out of four items correct for a skill area will indicate mastery of the skill. Thus, scores are reported in terms of the number of students who mastered the skill. The level of mastery is compared with that of previous years.

This description represents a fine test for a monitoring, original accountability test. However, it trails the curriculum; it does not lead the curriculum.

Another scenario would be a test like this. The test is focused on not only reading, language arts and mathematics but also science and social studies. Although basic skills are covered in this area, a significant part of the test covers critical thinking skills and interdisciplinary concepts. Teachers, curriculum specialists and administrators are involved in selecting learning objectives to be covered, but the emphasis for them is on defining objectives not commonly taught but which are highly important. They work together across the disciplines to define skills. The test developers still include items on the test for skills that are rather ambiguously defined; i.e., they do not measure only well defined skills.

The test is administered at the beginning and end of the school year to a sample of students at the grade level of interest. Students are selected who are being given different types of instruction related to the skills needing further refinement.

In the analysis, comparisons are made among the groups receiving different instruction. The items are carefully analyzed and discussed by the advisory group to gain a better understanding of what the students seemed to understand and what they did not. Corrections are made to the instruction received by various groups. Only on items commonly taught are comparisons made over time and reported to the public.

The Test Development Process

Although there are many decisions in the test development process that we could consider, we will focus on key decisions in five areas that inform the design and use of a state test:

1. The subject areas and learning objectives to be tested;

2. The test format;

3. Who to test;

4. Analysis methods; and

5. Reporting audiences.

My comments on these areas by no means present a complete picture of the issues involved. Rather, they simply highlight a few points that focus our attention on factors that can make a difference in whether a test plays a significant part in improving the curriculum and the environment of the school.

The chart in Table 1 which appears on the following page highlights the relationship between different testing orientations and the psychometric aspects of testing as organized by the steps in the test development process.

Identify Subject Areas and Learning Objectives. Noneducators often assume that any reading test, for example, can simply be taken off the shelf and used to measure reading. Not so. It is critical that careful thought be given to what specific skills (learning objectives) will be measured.

To date, many state assessment developers have done an excellent job of recognizing the need for careful delineation of skills. They have frequently used panels of teachers, community members, administrators and, at times, even students to help define the skills to be measured. This approach has been very positive in terms of helping to clarify education priorities. However, because the orientation has been primarily toward accountability rather than curriculum advancement, the skills actually tested have been limited to those frequently taught rather than those that should be taught but are not yet well understood in terms of how to teach the skill.

The implicit or explicit criteria for selecting the skills to measure often include whether the skill is widely taught and whether most students are likely to have acquired the skill by the grade level where testing is to occur. Such criteria are quite appropriate if the test is primarily for accountability. However, these criteria may be nearly the opposite of what is appropriate for a test to be used for curriculum advancement.

Table 1

Contrasting Testing Practices at Key Stages
in the Test Development Process

TEST DEVELOPMENT STEPS	TESTING PRACTICES	
	COMMON PRACTICE	CURRICULUM ADVANCEMENT PRACTICE
1. Identify Subject Area/ Learning Objectives	ones widely viewed as important and commonly taught	ones widely viewed as important but not commonly or effectively taught
2. Select Testing Format	relies heavily on well developed multiple choice items	includes use of open-ended questions scored using holistic, analytic, and primary-trait methods
3. Identify Whom to Test	all students at a given grade level or sampling only for efficiency	samples of different types of students in varying instructional contexts
4. Determining Analysis Methods	emphasize relationship to unalterable variables	analysis to determine relationship of performance to curricular issues and alterable variables
5. Identify Audiences for Reporting	public, school boards, teachers, administrators	also include: curriculum developers, study teams, researchers

Consider these two criteria a little further--first that of whether a skill is widely taught.

The skills and knowledge covered in a test oriented toward accountability should be teachable; i.e., good instructional materials should be readily available and most teachers in the given content area should have the training to teach the skills and knowledge. For example, most language arts teachers have been well trained in grammar and punctuation, and are very capable of teaching them. Many fine instructional materials are available. Unfortunately, we cannot say the same thing about writing (although improvements have been made in recent years). Many language arts teachers and other teachers have not been well trained to teach writing and the instructional materials are far more limited. Thus a test oriented toward curriculum advancement rather than accountability would be more appropriate in writing. A curriculum advancement test is premised on the idea that not all important subject areas and learning objectives are ready for widespread testing, either because many teachers are not well trained in the skills or because the teaching techniques or instructional materials are inadequately developed. Critical thinking and writing skills largely fall into this category.

Consider now the range of difficulty of skills measured. Accountability testing has moved us toward a limited range of skills--ones attainable by nearly all the students at a given grade level. If our goal, however, were clearly identified as curriculum advancement, a good share of the test would focus on skills that have not been acquired by many students. As we will discuss later, the items will be written and analyzed in a way that leads to understanding of why the skill has not been acquired rather than emphasizing which students have acquired which skills.

Currently we do not have a very good example of statewide testing that has served a curriculum advancement purpose--the testing of writing. In the mid 1970s, many districts, and then states, began to decry the poor quality of writing among their students. The response to such an outcry was to add writing to the subject areas tested through statewide assessment monitoring programs. One of the most important benefits of the widespread assessment was not the reporting of results for accountability purposes but rather involvement of many teachers in the scoring of those writing samples. As a result, they learned a tremendous amount about how to teach writing. This result of testing needs to become more than a by-product of testing. It needs to be seen as a key reason for testing--i.e., using testing as teacher training and thus a curriculum advancement tool.

Select Testing Format

Test items have two basic formats: multiple-choice items and open-ended items. As you know, multiple-choice items are by far the most common--and for some very good reasons. They are less-costly to score, especially with computers, and the scores are more reliable. Open-ended or essay items are much more costly to score. However, some higher level skills can only be measured well through open-ended items. Additionally, these open-ended items provide a special opportunity for curriculum advancement--the process of scoring actually can become a teacher-training activity. Most teachers involved in writing assessment said that scoring papers has taught them much about teaching writing.

Let me briefly describe the process of scoring writing samples. The description will underscore the usefulness of this activity for tests that serve a curriculum advancement purpose.

Students are asked to write several paragraphs in response to instructions such as: "Your school board is considering a policy that would support year-round school. Write a letter to the president of the school board indicating if you are for or against the policy. Present reasons for your position."

Three approaches to scoring are widely used.

Holistic scoring involves reading the student's paper for overall or "whole" impression. While specific traits such as organization, syntax, and originality undoubtedly influence a reader's judgement, none of these traits is directly, individually addressed. A reader makes a judgement in much the same way that he or she decides whether a novel or an essay in Time magazine is superior, mediocre, or slipshod.

Analytical scoring is a trait-by-trait analysis of a paper's merit. Individual traits--usually five or six--are those considered important to any piece of writing in any context. In other words, these traits would be important whether one were scoring an essay on migration of the Canadian geese or a letter of complaint to a local hardware store.

Primary trait scoring involves identifying the most important trait and scoring the paper on a scale related to that trait. Different traits are important to different pieces of writing. For example, in a set of instructions for assembling a bicycle, the most important trait may be clarity, for unless the instructions are clear and comprehensible, it is unlikely that the reader will be able to perform the task.

Teachers are trained in group sessions to understand the components of good writing and the particular scoring method to be used. They then practice scoring papers and engage in lively debate over how they have scored specific papers. They leave the session not only having accomplished the task of giving each paper a score, but perhaps more importantly, with a whole new understanding of what makes for quality writing and how they can better teach it.

It seems highly likely that we could transfer our experience in assessing writing to other areas such as critical thinking or higher order thinking skills. Those too are skills of which many teachers need more knowledge on how to teach and for which better instructional materials are needed. Tests of these skills could use open-ended items and analogs of the holistic, analytic or primary trait scoring methods used in writing. These scoring methods would provide the context for more teachers to learn the components of critical thinking and gain insights into better instructional approaches.

Statewide assessment tests are typically given to all students at a given grade level with the assumption that the results will be used by teachers to guide individual student instruction. Unfortunately, if the test simply covers commonly taught skills, teachers often already have this infor-

mation from district testing programs and/or the results are returned to them too late to be of much value.

Testing every student is extremely expensive. In a state oriented toward curriculum advancement, resources may well be directed toward a different selection of students to test. A sample of students may be tested to monitor trends over time but then the remaining resources may well be used to test students with different characteristics such as general ability levels or past achievement. Results could be carefully analyzed to identify links among curriculum content, instructional methods, and test results. Districts exploring new instructional approaches or teacher training could be special partners in the testing--not only in having their students tested but in the development of the test. Suppose you have discovered three districts in the state that are working on a new integrated curriculum focused on problem solving. Why not ask them to be partners with the state in identifying the critical skills, administering the test to a carefully selected sample of their students and assisting in the analysis and interpretation of the data?

Determine Analysis Methods. The possibilities for data analysis are nearly endless. The key to a good testing program is to have thoughtfully selected the analyses that will result in the most useful reports.

Typically, state assessment results are used to compare student achievement over time. Are this year's 4th graders performing better than last year's 4th graders? Do a group of students moving through the school system acquire certain skills as desired/expected?

One or both of two types of comparisons are typically used for state-testing--objective-referenced comparisons and norm-referenced comparisons. As you know, an objective-referenced test will focus on the specific skills a student has acquired. The public report typically states the number of students who mastered each skill.

In a norm-referenced test, the emphasis is on comparing one student's performance to that of other students. This type of test would generally have "norms" for, say, fourth graders--i.e., the test publisher would provide tables that allow you to show how well a given student did compared to a representative sample of fourth graders tested by the publisher. The score would be reported in units such as percentiles. When we say, "Johnny was at the 60th percentile," we mean that he did better than 60% of the fourth graders tested by the publisher. While this is useful information, it does not reveal information on specific skills.

If a standardized test is used, it is becoming increasingly common for both types of scores to be available. Using either type of score, analyses are usually done that show trends over time for groups tested.

Methods for analyzing a test designed for curriculum advancement purposes are likely to be noticeably broader than those just described, both because different groups of students would be tested and different information would be desired. Consider the situation that was described earlier where a test of critical thinking skills was being developed by the assessment unit in the state education agency working closely with several content specialists in mathematics, science, social studies and language

arts who are looking across the content areas to gain a better understanding of what the components of critical thinking are. The analyses of most interest to them could likely involve measuring the performance of groups of students before and after they had received different instructions. The item analysis would be studied to see both how the items could be improved as well as what could be learned about the various instructional methods used with the students tested. For curriculum advancement purposes, analysis methods need to promote a better understanding of how skills can be discreetly identified, taught and measured. Once the skills are defined in a teachable fashion, not just a measurable fashion, we are well on the way to identifying skills to be measured in subsequent testing for monitoring and accountability purposes.

Identify Audiences for Reporting. Although this is the last topic being discussed and the last step in the assessment process, it should be one of the very first topics considered when developing a testing program. If your purpose is accountability, the most likely audiences will be the public, school boards, administrators, teachers and the students. Each audience has a somewhat different interest. It is critical that their interests are carefully thought through before the testing system is established. What is to be reported will significantly shape each of the decisions we have already discussed.

The audiences for a curriculum advancement test are likely to be different from those for a monitoring test. Curriculum developers, administrators, study teams, researchers and teachers are the more likely audiences. However, it may be very useful to report the general findings and their implications for shaping the curriculum to the public, school boards, and others as well.

Using Testing Choices To Promote Excellence in Schools

We have now looked at the process of test development discussing two different ways the process can be structured. I hope these comments have highlighted a few of the critical test development choices and how these choices relate to test uses as well as technical feasibility. Consider now how some of the choices discussed relate to whether tests will promote or hinder positive characteristics in schools.

Recall that I mentioned several characteristics of effective schools:

1. high expectations for all students;

2. instruction guided by a clearly defined and focused curriculum;

3. close monitoring of student progress; and

4. an atmosphere of collegiality and improvement.

Can a state promote these characteristics in schools using a prudent combination of monitoring, accountability and curriculum advancement testing?

First, high expectations for all students and a clearly focused curriculum can be promoted with a combination of (1) test components that monitor achievement of a variety of skills that are widely viewed as important and

teachable and (2) other components that demonstrate that the state is seeking to promote better understanding and teaching of critically important skills that have not been widely taught. By using testing as a tool to stimulate creative thinking about how to deal with these little-taught skills, the state is acknowledging that further development work needs to be done before all schools can be expected to effectively teach these skills to all students and that it is interested in promoting new ideas in these areas.

Involving teachers and administrators in the development and use of creative testing efforts can be very effective in keeping educators clear on their educational purposes. We are all too aware of how frequently curriculum guides sit unused on classroom shelves until tests begin to measure those skills. It would probably be better if, through good school management and clear understanding among teachers of educational priorities, the appropriate skills were well taught and learned without the threat of a test. But realistically we know that a test can be a very useful tool to emphasize those priorities, especially when test results will be made public.

Concerning the monitoring of student progress, state tests alone are not sufficient for monitoring student progress. However, they can be very useful in determining, at key decision points in students' lives, if they are acquiring major skills and knowledge. State and local educators working together to test and monitor progress on important but poorly taught or understood skills can be an especially effective way to advance the curriculum at both state and local levels.

Distinguishing different testing purposes can be especially helpful in promoting an atmosphere of collegiality and improvement. Rather than using accountability testing to hold teachers responsible for teaching skills for which they are not sufficiently prepared, an atmosphere of collegiality and improvement can be promoted by separating skills for accountability testing from those that need to be covered in curriculum advancement testing. Such distinctions allow all parties concerned to agree on a set of skills that are immediately legitimate for coverage in monitoring or accountability testing and also to work out testing that will serve a curriculum advancement purpose better for some important but little-taught skill.

Summary

Testing is a powerful force in education, one that is increasingly being used to shape the process. We have great responsibility to ensure that it is a force that truly benefits children and not a force that feeds bureaucracy and dampens creativity and enthusiasm. By clearly understanding the characteristics of effective schools, by knowing which characteristics we wish to promote, and by thinking carefully about our purpose for testing, we can make testing more useful, more effective, and more integral to the process of school improvement. "We must take the current when it serves, or lose our ventures...."

Educational Implications For State-Mandated Student Assessment on the Classroom System

by

David S. Lane, Jr.

State-mandated student assessment will have important philosophical, psychometrical, financial, and political ramifications; however, it must be recognized that all of these areas of implication converge with their impact on the educational arena--specifically the school with its individual classrooms and its individual students. There are perhaps as many approaches to discussing the educational implications of statewide assessment programs. It seems imperative that any of these approaches involve the many facets of the educational setting including the child, the instructional materials, the instructional outcomes, and the instructional measures. One approach which seems particularly germaine is looking at the educational implications of state-mandated student assessment through a systems approach.

A system can be defined as a set of individual yet interdependent components functioning separately yet conjunctively (eg., Kaufman, 1972). Furthermore, a system implies that altering one aspect of the system necessarily alters other components of the system. In effect, changing one part of a system causes the whole system to change with the net result being a "new" system.

Using this definition as a background, it is a matter of application to see that the school, the classroom, and, for that matter, the individual child are all examples of systems. For instance, it is apparent that when one teacher of a classroom changes a procedure, the school as a whole becomes a little different. Certainly, when one child is absent, the classroom dynamics are altered and when a substitute teacher is present, classroom procedures are never quite the same. Each of these systems is affected and changes when its components are altered.

For the topic at hand, three particular systems--the school, the classroom, and the child--will necessarily be affected by state-mandated assessment. Of the three systems, some might argue that the primary educational implications of the state-mandated assessment will be felt at the school system level since the school system encompasses the other two. In fact, much of the literature addresses the educational implications at the school system level. Realistically, however, it seems that the most practical effects will be at the classroom level. It is at the classroom level where adjustments will have to be made to accommodate the state-mandated assessment process. These judgements will include everything from simply allowing time to administer the assessment instruction to changing teaching methods in response to results. Moreover, if one adopts the systems approach, discussing the effects of assessment on the individual classroom will necessarily involve the school system and the systems of the individual student since the classroom is an integral aspect of both of these systems. Thus, this paper will address the educational

implications of state-mandated student assessment on the classroom "system."

One way to construe the classroom system and to discuss the effects of state-mandated student assessment is to consider the classroom's primary purpose--instruction. Instruction has been defined as "a systematic process in which every component is crucial to successful learning" (Dick & Carey, 1985, p. 2). It seems clear that state-mandated student assessment will affect the instructional process (eg., Perkins, 1982). Consequently, it seems obvious that an important aspect of instruction is the planning, the design, of the instruction. If the planning and design of instruction will be affected by mandated assessment, then one way to investigate these effects is within the realm of instructional design.

Dick & Carey (1985) present a procedural approach for investigating and designing instructional systems which includes preparation, implementation, evaluation, and revision. These four phases are then broken into the specific components of the system. Examples of components in the instructional process include goals, characteristics of the learners, performance outcomes, and instructional content. Each of these components in the instructional design process will be affected by a state-mandated assessment program, thus having an impact upon the classroom, its teacher, and its students.

In instructional design, the first component is determining the goals, based on needs, of a classroom. This component is then used in formulating instructional evaluation measures, determining instructional content based on characteristics of students, and setting instructional strategies. In a sense, setting goals and objectives becomes the primary building block in the instructional design; consequently, setting goals and objectives will be the main point of discussion for this paper.

Many critics (eg., Popham & Lindheim, 1981) stress the need for mandated assessment, especially minimum competency tests, to be linked to the goals and objectives of the curriculum. McLung (1979) calls this "curricular validity." Instructional design procedures suggest that curricular validity is developed by formulating assessment procedures and instruments after goals and objectives have been set. In this sense, it is imperative that this first component of instructional design, setting goals and objectives, be considered by the state-mandated student assessment process.

It seems logical that if a test is to measure or assess a student's competency in various curricular areas, then that measurement instrument must be linked to the objectives and purposes that the particular curricular areas address. McLung (1978) complicates the issue by contrasting curricular validity, where test objectives are matched with course objectives, with "instructional validity," which he defines as an actual measure of what is being taught in the classroom.

Theoretically, in the instructional design process, instructional and curricular validity are the same. Practically, the two types of validity are probably too often very different. For instance, many goals of the classroom emanate from the state level and are aimed at the "typical" classroom. Sometimes these goals are necessarily broader to encompass all of the "typical classrooms." This broadness allows for interpretation at the

classroom level but also allows for incongruency between curricular and instructional validity and makes the mismatch between curricular and instructional validity a viable reality. Consequently, it is possible that an assessment measure might have curricular validity but not instructional validity (Popham & Lindheim, 1981). This is a very real concern for the individual classroom where teachers often modify curriculum objectives to meet the individual and unique instructional needs of their particular classrooms.

One way to avoid "teaching for the test" and to insure the match of curricular and instructional validity is to guarantee that the objectives for the student assessment instrument are true reflections of instructional needs and practices. Needs assessment is one method of ascertaining appropriate goals in the instructional design process. Needs assessment is a process where input is obtained from a variety of sources such as parents, teachers, students, colleges, curriculum experts, etc. These sources are used to assess what the perceived versus actual needs of an instructional setting are and should be. This information is then used to write the goals which will better reflect what that instructional setting actually needs rather than just a conglomeration of ungrounded assertions.

Basing instructional goals on a true needs assessment process then requires that instruments be selected or developed to measure these goals, and then instructional content be selected which will help students reach the objectives identified by the needs assessment and measured on the assessment instruction. This process would weaken the argument that teachers would simply "teach for the test" as then teachers would not be teaching for the test but the test would be measuring the attainment of the objectives for which test items have been written. The results are different if, on the other hand, the objectives of the assessment instrument do not match the actual objectives for what is actually taught. If the latter occurs, a more serious outcome is that teachers, because of perceived or real pressure, might be "forced" to change what is taught to match specific items on the test; i.e. to teach for the test.

Appropriate instructional design procedures dictate that evaluation measure whether goals and needs have been obtained (eg., Gagne & Briggs, 1979); however, the goals and objectives are formulated at the beginning of the design process; the evaluation is formulated afterwards. One concern is not to let state-mandated assessment reverse this process.

As the systems approach to the instructional process requires, changes in the goals and objectives will automatically alter the other components in the process. Three other components will be discussed briefly to present an idea of how changing the goals and objectives of a classroom system affects other classroom system components. Students in the classroom is a second component in the systems model of instruction. As discussed above, the individual student is, in one sense, a system of its own. Yet as each child "system" is affected so is the classroom system. Consequently, for purposes of this paper, the student will be addressed as one vital component of the instructional process in the classroom system which will be affected by state-mandated assessment procedures.

There are several points to consider when discussing the effects of mandated assessment on the individual child. One of these points is the

impact of this type of assessment on minority groups, exceptional children, gifted children, and/or socially and economically deprived children. It cannot be doubted that such an impact exists on practical, legal, and theoretical issues (eg., Citron, 1982; Linn, Madaus, & Pedulla, 1982; Ross, 1982). Linn, Madaus, & Pedulla (1982) state that failure rates on minimum competency tests for certain minority groups and some groups of handicapped students are much higher than for white, nonhandicapped students. The authors argue that minimum competency tests compound the already present injury of discrimination, poor placement, and inadequate education for these students. Ross (1982), on a more positive note, suggests that minimum competency tests could serve to foster integration of all children into the educational mainstream; she does, however, indicate that this integration will not be an easy task.

What these arguments suggest is that state-mandated assessment will affect "special" students and that considerations of practical and theoretical issues such as cut-off scores, who should take the test, and use of the scores must be addressed by the users of the test results. The instructional design process often uses preassessment of learners' ability based on the objectives. Students proficient in the various skills are then placed on different content tracks than those students deficient in the skills.

Just as these "special" students will be affected by state-mandated assessment procedures, so to will every other child in the classroom. One effect might be increased anxiety. The relationship between anxiety and test performance is well documented by the Yerkes-Dodson Curve. This relationship suggests that optimal performance is obtained when students are in a moderate state of anxiety, suggesting that, since the teacher wants optimal performance, he/she should attempt to attain "moderate" anxiety in the students. This, obviously, is easier said than done as teachers may overestimate anxiety levels and may cause increased and unnecessary anxiety in some children. The potential for increased anxiety would seem to be present regardless of the purpose of the test but would seem to be especially present with state-mandated tests where certain scores were required for grade promotion or graduation.

Another, more subtle, effect on the child might come about by a perceived shift in classroom structure. This shift might occur as teachers respond to administrative and public pressure to raise test performance. Such pressure, whether overt or covert, would obviously affect the classroom as teachers would alter their methods, materials, etc. to better test performance. Such altering would be a positive change if the primary goals of the classroom coincided with, and were supported by, the change; i.e. a learning-centered approach. Such altering could be adversive if the changes were simply to improve a score on the test and were not cohesively tied to the overall instructional purposes; i.e. a performance-centered approach. Even more damage might occur if students were caught in the middle of this learning- versus performance-centered controversy. This is not to say that the two approaches cannot coexist, it is only to suggest that to make the two approaches coexist takes integration, planning, and thought.

A third component of the systems approach affected by mandated assessment would be the various classroom tests used by the teacher. Optimally, this

component is linked to the original instructional goals set at the beginning of instruction. Good instructional development procedures dictate that in-class performance measures reflect overall instructional goals. Thus, as argued above, if the instructional validity matches the curricular validity, in-class performance measures will support performance on the assessment instruments. Conversely, if the instructional-curricular validity match is absent, in-class measures could be redundant, unrelated, or inappropriate.

One last system component to be discussed is instructional content which includes the material presented to the student, including material, media, and method. Typically, one purpose of mandated assessment is to look at student performance in content areas. As introduced above, in a typical instructional design sequence content is developed after assessment procedures are developed to measure performance on previously developed goals and objectives. What the above arguments suggest is that planning must ensure that what is being taught reflects actual, as opposed to perceived, goals and needs, and that any state-mandated test measure these actual goals and objectives. If classroom content is linked directly to the objectives for a classroom through the assessment instrument, then performance on the instrument would be a more reliable indication of student progress for those objectives and a more reliable indication of the effectiveness of the methods being used.

If one accepts that the instructional content that is taught should tie in with the instructional goals and objectives, a seemingly universal assumption in education, then it is not difficult to understand how state-mandated assessment will affect the classroom instructional content. For if state-mandated assessment measures instructional goals, then the way the goals are taught will be related to the mandated tests since the interdependency of components within the classroom system obligates systemwide change when one part of the instructional process changes. Teachers will have to, and will want to, select materials, media, content, and methods which reflect the instructional goals and which allow their students to perform optimally on the assessment measures. These changes can be beneficial to the students if the altered instructional content supports the goals of the classroom and contributes to the instructional validity of the state-mandated assessment process.

This paper has explored some of the issues surrounding state-mandated student assessment as it relates to the instructional process in the classroom setting. Specifically, a systems approach was introduced as a method system. Through this brief discussion of some of the educational implications, it seems clear that any state-mandated program will have a direct effect on the classroom. In this sense, state-mandated student assessment is not unique. What does make state-mandated student assessment unique is its potential for profound influence on the goals and objectives of the classroom which will then affect students, teachers, and content.

References

Citron, C. H. (1981). Competency testing: Emerging principles. Educational Measurement: Issues and Practice, Winter, 10-11.

Dick, W. & Carey, L. (1985). The Systematic Design of Instruction (2nd Ed.). Glenview, IL: Scott, Foresman and Company.

Gagne, R. M. & Briggs, L. J. (1979). Principles of Instructional Design (2nd Ed.). New York: Holt, Rinehart and Winston.

Kaufmann, R. A. (1972). Educational System Planning. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Linn, R. L., Madaus, G.F., & Pedulla, J. J. (1982). Minimum competency testing: Cautions on the state of the art. American Journal of Education, November, 1-35.

McLung, M. S. (1978). Developing proficiency programs in California public schools: Some legal implications and a suggested implementation schedule. Sacremento, CA: California State Department of Education.

McLung, M. S. (1979). Competency testing programs: Legal and educational issues. Fordham Law Review, 47, 651.

Perkins, M. R. (1982). Minimum competency testing: What? Why? Why not? Educational Measurement, Winter, 5-9, 26.

Popham, W. J. & Lindheim, E. (1981). Implications of a landmark ruling on Florida's minimum competency test. Phi Delta Kappan, September, 18-20.

Ross, D. (1982). Competency based education: Understanding a political movement. The Educational Forum, Summer, 483-490.

State-Mandated Student Assessment Programs: Educational Implications

by

David Yellin

The movement towards state-mandated student assessment programs is growing. And with that growth has come attendant problems, not the least of which is a confusion of terms. Thus, let us begin with a few definitions in common. Statewide student assessment programs are generally referred to under the rubric of Minimum Competency Tests (MCT). The latter are administered to all students in a state, with reading, math and language arts being the most commonly tested areas. The results of the MCT's are then used in one of two ways: tests are administered at selected grade levels, say 4, 7 and 11, to justify promotion and retention or the tests can be used as an exit requirement from secondary school and a condition for receiving the high school diploma. A few states use MCT's for both purposes (Haney and Madaus, 1978).

Contrary to belief, statewide student assessment is not new. In fact, it has a long and controversial history. In the 1840's, Boston's Superintendent of Schools ordered that a passing grade on a common exam be a prerequisite for graduation from high school. The state of New York began administering its Regents Exams in 1877 as part of the high school graduation requirements; to graduate from New York high school with an academic diploma a student must acquire a certain number of Regents credits by passing standardized tests in reading, math, language arts, science and social studies. This is in addition to receiving passing grades in the subject class.

Currently, 17 states have laws requiring some form of MCT for either grade promotion or high school graduation; another dozen states have districtwide competency tests and are considering state legislation making the exams mandatory for all students. Florida's MCT has been used as a model for many other states but has also been challenged in the courts. The tests have sparked controversy elsewhere, too. In Connecticut, a proposed 7th grade reading equivalent score for high school graduation caused the school board members to complain that the citizens of Connecticut were "paying for 12 years of schooling and getting only seven years" (Nathan and Jennings, 1978). Similar complaints were heard in New York state where a 9th grade equivalency score was being used. There was bitter debate in the California state legislature over the Hart Bill which requires minimum standards for high school graduation. In fact, no region of the nation is immune from the controversy over testing.

In Oklahoma, the legislature recently passed H.B. 1466 which charges the State Department of Education to draft a plan to implement an Oklahoma School Testing Program. The proposal is to test students in grades 3, 7, and 10 in math, reading, communications, science, citizenship and language arts (House Bill 1466, 1985). However, it is not yet clear how the test scores will be used, whether functional literacy skills will be required

or who will prepare the actual exams. The questions facing Oklahoma are indicative of the problems associated with Minimum Competency Testing in general.

A review of the literature reveals four main problems which are briefly summarized here as they relate to an overall understanding of the MCT's impact on education. First, it is still not clear what constitutes the minimum competencies. Is the ability to add and subtract fractions a minimum competency? Or should the ability to balance a checkbook be the minimum criteria? Second is the touchy question of how shall the cut-off scores for pass and fail grades be determined? Even the experts in this area admit that there are no reliable statistical techniques for setting the cut-off score; instead it remains a subjective judgement that will vary from one test maker to another. The question of the validity of the MCT presents a third problem. Does a score on the MCT truly represent a student's ability in a subject area and does it really predict success beyond school? Again the experts are divided. And finally another matter associated with Minimum Competency Tests is the impact they will have on minority populations. In Florida, for example, a large percentage of black students failed the state's Minimum Competency Test requirement for the high school diploma. In the case of Debra P. vs. Turlington (1983), a class action suit by Florida's black students was brought against the Superintendent of Schools, claiming that the test was unfair, did not test what the students had been taught, and discriminated against blacks. Eventually the U.S. Fifth Circuit Court of Appeals ruled that although the state's test was legal it was indeed biased, did not test what the Florida schools actually taught and should be modified accordingly.

Despite problems, controversies and even court cases, it is clear that Minimum Competency Testing is a movement that is here to stay. It is, in large part, a reflection of the times in which we live. The political shift to the right, the growing strength of conservative religious groups and the underlying philosophy of the back-to-basics movement all have influenced the spread of MCT's across the nation. Under such circumstances educators are now grappling with the significance of this testing movement for the schools and students of the future. The main educational implications of the MCT movement center around the issues of control, displacement, achievement, and distraction.

Beneath the esoteric language of test manuals and legislative mandates the central issue underlying MCT's is that of control. Who will make the decisions that affect what goes on inside our school buildings? Who will be the educational leaders of the future? As the influence of statewide assessment programs spreads and the parallel press attention they receive increases, it appears that the tests themselves and those who prepare the tests will more and more shape the future of education in America. Control of education will rest in the hands of the test makers and beyond the influence of teachers and administrators. Stanford University's Elliot Eisner describes the situation this way:

> Communities led to believe that the quality of education is represented by the reading and math scores students received come to demand that those areas of the curriculum be given the highest priority. When this happens, teachers begin to define their own priorities in terms of test performance...test scores

> function as one of the most powerful controls on the character of educational practice. (Eisner, 1979, p. 2)

This notion of control from without, by noneducators in many cases, will increase as the influence of test scores grows. Students, and their parents, who believe that what really counts is the score on the test not performance in class, will turn towards after-school "academies" which coach students on how to score high on specific tests. Such private test-training institutions, once confined to prepping students for S.A.T. college entrance exams, will proliferate as more and more states adopt MCT's. Thus the private entrepreneurs operating such academies may in the future exercise as much influence over the nature of education as the professional teacher and administrator.

As the control over education shifts into the hands of standardized test makers, the managers of testing academies and other noneducators, teacher and school performance will be measured against standards set from afar. Successful schools will be those which meet the standards (e.g., have the highest scores on the tests). Schools which fail to meet the standards may lose state or local financial support; teachers may lose their jobs. And the complex human process of education will have been reduced to the simplistic notion of a production line turning out perfectly made "widgets" and rejecting the flawed ones.

Closely related to the issue of control is the concept of educational displacement. By displacement is meant the devaluing and diminution of one educational area in favor of another. Since the all important MCT's emphasize certain subjects over others, and teachers would teach accordingly, then those subjects not tested will be displaced or perhaps eliminated from the curriculum. The back-to-basics movement would teach reading, writing, and arithmetic exclusively. And the minimum competency advocates would test only those subjects. Again, Eisner puts the situation succinctly:

> Schools whose quality is judged in terms of the test performance of their pupils tend to emphasize what the tests test. This means that the fields not tested are considered less important than those that are....(1979, pp. 4-5)

It is thus easy to imagine a future for education in which primary grade children learn the 3Rs and little else; the role of play, the arts and social skills, since they are not tested, would be ignored. In the elementary grades, social studies instruction should not be relegated to the spare moments of the day when children need a break from the real work of school. Art and music will be virtually nonexistent or assigned to after-school programs.

A third issue relates to academic achievement and the overall quality of education for the future. Advocates of MCT's have always assumed that the test will somehow improve the academic achievement of students and the quality of education in general. The opposite may instead be the case: a deterioration in both achievement and quality. This is because Minimum Competency Tests are almost exclusively skills oriented. In designing the tests, broad general knowledge and subject areas are ignored in favor of subject matter which lends itself to multiple-choice tests. Usually this

means testing low-level cognitive abilities. And the higher level thought processes, since they would not be tested, would eventually not be taught. Thus the overall academic achievement of students and the general quality of education they receive would be diminished. Two examples from the current state MCT's support this argument.

Experts in the area of reading instruction have long argued that the goal of reading should be to teach students to comprehend and analyze printed material. Yet at the 1981 National Institute of Education hearings in Washington, D.C., Roger Farr, reading authority from Indiana University, showed that the majority of MCT's in use measured only literal comprehension and word recognition in isolation, two low-level cognitive skills (Linn, Madaus, Pedulla, 1982).

In the testing of writing skills the MCT situation is actually ludicrous. Writing authorities know that writing ability can only be demonstrated by having students do real writing exercises, not by filling in blanks, labeling parts of speech or identifying grammatical errors (Graves, 1983). Yet the latter is all that the MCT's require of students. In other words, the MCT's purport to measure students' writing abilities without requiring them to write!

A fourth and final issue which should concern us is the problem of distraction caused by the Minimum Competency Test movement. The uproar over and publicity given to the various MCT programs around the country have distracted educational reformers away from the significant social and economic issues that impact on the schools and affect the quality of education. A few statistics should suggest the challenges facing education in the future:

1. One in five children in America live below the poverty line.
2. One in five children born in 1982 were born out of wedlock.
3. Fifty-three percent of school-age children come from homes where both parents work. (Feistritzer, 1985).
4. More than 1.3 million students receive instruction in their native non-English tongue in America's schools. (Bowen, 1985).

Clearly the children in school today have different backgrounds and needs than those of children in the past. And those real differences pose problems for the school and society that MCT's do not even begin to address, but may indeed blind us to. The Florida court case (Debra P. vs. Turlington, 1983) demonstrated that neither the schools nor the testing programs can afford to ignore the population they serve. The issues of poverty, children without parents, and non-English speaking students directly affect the nature of the classroom and the institution of schooling. At a time when more reform and more support is needed, the Minimum Competency Test movement threatens to divert educators' energies and scarce educational dollars away from the real problems. As the MCT movement gains momentum, educators may see only raising test scores, rather than improving the process and quality of life in school, as the significant goal of education.

The four issues of control, displacement, achievement and distraction are only some of the implications for education suggested by the state mandated student assessment programs. The trend towards more and more

state-wide standardized testing continues unabated. Earlier in the 1980's, a debate was sparked by the publication of A Nation At Risk. Perhaps the Minimum Competency Test movement, with the strong feelings it evokes on both sides of the issue, will spark a new debate over the future of education in America.

References

Bowen, Ezra. For Learning or Ethnic Pride. Time Magazine, July 8, 1985.

Eisner, Elliot. The Educational Imagination. New York: MacMillian Pub. Co., 1979.

Feistritzer, Emily. Commentary News Bulletin. NCEI Report, Washington, D.C., February 11, 1985.

Graves, Donald. Writing: Teachers and Children at Work. Exeter, New Hampshire: Heinemann Ed. Books, 1983.

Haney, Walt and George Madaus. Making Sense of the Competency Testing Movement. Harvard Educational Review, Vol. 48, No. 4, November 1978.

House Bill 1466. Oklahoma State Legislature, Oklahoma City, OK, 1985.

Linn, Robert, George Madaus and Joseph Pedulla. Minimum Competency Testing: Cautions on The State of the Art. American Journal of Education, U. of Chicago Press, November 1982.

Nathan, James and W. Jennings. Educational Bait-and-Switch. Phi Delta Kappan, 59, pp. 621-24, 1978.

State-Mandated Student Assessment Programs: Political/Financial Implications

by

Joseph R. Weaver

The history of state-mandated student assessment programs in the United States verifies that the question of student assessment through legislated testing programs is a volatile issue. Although proficiency testing is not a new phenomenon, the tremendous growth in the number of statewide testing programs has occurred within the last two decades. Between 1963 and 1974, 30 states enacted legislation pertaining to educational accountability, and by the late 1970's most states had either statutes or administrative regulations mandating procedures to ensure that schools produce specified results in terms of pupil achievement.[1]

The minimum competency testing movement has been a natural outgrowth of two societal trends: 1) schools have received increasing criticism for graduating students who are not adequately prepared for citizenship and employment; and 2) taxpayers have become more vocal in demanding accountability for the use of tax dollars to support public schooling.[2]

Debate about state-mandated testing programs, as today's agenda attests, usually is focused on legal, psychometric, philosophical and educational issues. Since any of these categories of issues is potentially a political issue, some of my discussion may cross the boundaries implied by the listing of topics for the concurrent sessions.

Because of the breadth of the topic and the limited time allotted for this session, I will limit my remarks to global issues relative to the political and financial implications of state-mandated student assessment programs. My discussion will be centered around three areas: 1) issues that have political/financial implications; 2) groups that emerge to apply political pressure on decision makers; and 3) financial implications.

One of the major issues relates to the purposes of testing programs. A number of possible purposes have been identified:

> Accountability; i.e., as a means of assuring taxpayers that money appropriated for education is well spent;
>
> Needs assessment; i.e., determining the learning needs of students, diagnosis and remediation;
>
> Program improvement; i.e., curriculum and program development, program research, staff development, and instructional improvement;
>
> Promotion/graduation criteria, grade/level placement;

Comparison of norms; i.e., state/national, district/state, district/national, district comparison (grouped by categories of like districts), and school to school comparison within districts;

State/Federal program eligibility; and

Teacher evaluation on basis of student performances on tests.

Variables that should be considered in comparing school districts are:

Finances--per pupil expenditures, community tax base, per capita income;

Number of students in each district;

Average number of years of schooling of adult population in the district;

Population (size) of each district; and

Racial/ethnic composition of each district.

The National Education Association has stated that the major purposes of student evaluation should be (a) to help students choose more wisely among options and to learn better, and (b) to assist teachers in selecting the most appropriate learning experiences for individual students. The use of evaluation for categorizing and tracking students and for denying them educational opportunities should be deemphasized in favor of constructive, diagnostic, helping purposes.[3]

Opponents of student assessment programs will argue that the purposes of testing can be accomplished through a variety of alternatives. Settling the argument between proponents and opponents of student assessment programs may require a political process.

Another major issue with political implications is related to test characteristics. That is, what type of tests should be implemented--minimum competency, standardized achievement, criterion-referenced, norm-referenced? The answer to the question depends in large measure on the purposes of the testing program. For example, if the primary purposes of the testing program are diagnosis and remediation, curriculum and program development, and instructional movement, then a criterion-referenced test would be the most appropriate type. On the other hand, if the primary purpose is to compare the achievement level of children in a state to the national norm, then a norm-referenced test would be the better choice.

The possibility of adverse impact on minorities is another issue having political as well as legal implications. The court decisions in the Debra P. vs. Turlington case in Florida have been instructive for test developers. States have a moral as well as a legal obligation to ensure that tests used in statewide assessment programs are free from cultural bias and that all appropriate steps are taken to ensure fairness to all examinees in testing procedures.

The National School Public Relations Association has identified several factors that have created concerns which have political implications. These factors are:

1. Generation and publication of rankings (percentiles) of school and district scores;

2. Use of socioeconomic data gathered from students (even from anonymous surveys);

3. Use of student attitude surveys (attitude toward self, toward others, toward school, etc);

4. Use of assessment data to allocate funds to selected schools or districts;

5. Fear on the part of teachers that performance is being judged by students' scores;

6. Accusations made by minority groups that the tests are biased; and

7. Accusations made by districts that the tests are not valid (or are not reliable).[4]

The availability of statewide test data raises the issue of control of school policy; i.e., local vs. state control. The whole issue of controlability/noncontrolability; i.e., which variables can be controlled at all and which can be or should be controlled by a state agency or decision-making body, has major political implications. Prioritizing the variables that can be changed or controlled by a state agency becomes an important political as well as educational and philosophical issue.[5]

If better test performance is an objective of the decision-maker, what are the controllable variables and what is their relationship to test performance? What courses of action will be chosen to meet the objectives? This is partly a political process. Financial support to implement the courses of action must then be provided.

Additional issues that have political as well as educational implications are: Who determines the purposes of the assessment program? What test will be used? Who sets the standards? How will they be established? How can fairness for all students be insured? Should sanctions be applied for failure or poor performance? If so, what sanctions? What exemptions, if any, should be allowed? What options should be made available for those who fail? How can test performance be explained? That is, why are the test results as they are? Determination of reasons or explanations for student performance in the absence of other important data can lead to inappropriate conclusions.

Sources of political pressure related to implementation of a state-mandated assessment program may be inferred from the types of issues discussed above. These sources include teacher groups who view student testing as a method to evaluate teachers and eliminate those who do not measure up to established or implied standards, groups of educators who

oppose student assessment programs on philosophical grounds, and minorities who oppose student assessment because of a fear of cultural bias.

Other groups that may seek to influence decision-makers through the political process are administrators and teachers who fear that test data may be used to compare school districts strictly on the basis of student test performance. Special education teachers may want special students to be exempt from testing requirements or may want them included in the testing, depending on the type of tests used and other factors. They may use political action to achieve their goals.

For almost any philosophical issue relevant to student assessment there is potentially a group that will bring political pressure to bear on decision-makers in support of its position. The responsibility of decision-makers, then, is to choose appropriate courses of action based on the best available data and sound logic.

The assumption has been that statistics, including test data, help create a better information base to use in making educational decisions, and in funding improved educational programs. Student assessment programs impact school financing in many more ways than the funding of test development (or adoption) and administration. These are the obvious costs of such programs.

Additional costs that may be generated by implementing a statewide student assessment program include costs for curriculum and instructional development programs. Textbooks and other curriculum materials need to be analyzed to insure appropriate match with the test objectives. Remediation programs may be needed to assist poor test performers.

Funds may be targeted for staff development related to the use of test results. For example, in-service may be needed dealing with test score interpretation and prescriptive instruction. School improvement programs tied to state testing programs could require additional funding at the local level. Another potential cost at the local level is that incurred by adjusting local testing programs to conform to a state testing program.

Once test performance data is available it will be necessary to identify those environmental factors that are contributors to high performance and those that contribute to low performance. When these are identified, resources must be secured to equalize performance across districts. If test performance is an indicator of educational quality, then test data must be used as a basis for decisions about funding priorities.

In summary, it is clear that there are many political and financial implications of a state-mandated student assessment program. The political implications emerge from the philosophical, legal, psychometric and educational issues that surround the implementation of state-mandated student assessment programs. Each philosophical issue will attract a support group that will apply political pressure on decision-makers to achieve desired results. The availability of test results provides decision-makers with data upon which plans may be formulated to improve instruction, promote effective schools and enhance student learning.

References

[1] Phyllis Hawthorne, Legislation by the States: Accountability and Assessment in Education (Denver, Colorado: Cooperative Accountability Project, 1984), p. 3.

[2] Manth M. McCarthy, "Minimum Competency Testing for Students: Educational and Legal Issues" Educational Horizons, Spring 1983, p. 103.

[3] Frances Quinto and Bernard McKenna, Alternatives to Standardized Testing, National Education Association, 1977, p. 7.

[4] Releasing Test Scores: Educational Assessment Porgrams, How to Tell the Public, National School Public Relations Association, 1976, p. 8.

[5] Preston C. Kronkosky, Applied State Level Educational Decision-Making: A Proposed Model, Unpublished Dissertation, The University of Texas, Austin, 1970.

State-Mandated Student Assessment Programs: Finance

by

William E. Camp

The Oklahoma state legislature mandated state assessments of student achievement for the 1985-86 school year. As the legislative session ended late in the summer break, school districts and the state department were left with little time to properly implement this requirement. The resulting rush by the state department to design and bid this program left little time for communication with school districts.

Individuals responsible for district programs are looking for answers concerning the intent and procedures for the program. It is clear that the programs are designed to provide educational and political leaders information to sustain public and financial support for educational improvement. Part of the emphasis must come from the "band wagon" effect of recent educational studies related to excellence in education. Over 50% of the states have implemented these programs. As an example, in the Southern region Oklahoma has added a norm-referenced testing program; Mississippi, Kentucky, and Tennessee are starting competency testing; and South Carolina and Texas have mandated basic skills testing for high school graduation.[1] Other goals are prominently mentioned in various documents relating to the programs.

Instructional improvement is an important goal mentioned in the initial proposal to the legislature. "Effective assessment of learning relies upon multiple methods of measuring student achievement. Test results could provide data upon which decisions can be based regarding curricular and instructional improvements."[2] The RFP also focuses on this issue: "Bidders are requested to report these data in the must useable ways (e.g., through use of a grid or other such graphic means to facilitate application of these results to the improvement of instruction for individual students as well as for program development and improvement."[3]

Remediation is another important projected outcome under instructional improvement. The proposal to the legislature states, "The results gained from such a program could be used to determine when remediation or enrichment is needed."[4] The legislative act states, "The State Board of Education shall seek to ensure that data yielded from the test is utilized at the school district level to prescribe skill reinforcement and/or remediation by requiring school districts to develop and implement a specific program of improvement based on the test results."[5]

Accountability is another important goal both stated and alluded to in various documents and conversations. The stated purpose is "...the testing program shall be to afford a component for use along with other pertinent data in evaluating the effectiveness of public schools as shown by the competence and progress of pupils in basic skills."[6] It is evident that comparisons of programs will be made. It is not clear how these comparisons will be used. Dr. John Folks states that the reports will be

used to help his office make decisions for future support for Oklahoma schools. He will look at the state as a whole and not on an individual district basis. The examination will be prescriptive in nature based on what needs to be done for the students to improve delivery of educational services.

The cost of this program is very important in future decisions. A statewide survey in 1984 showed that of the 457 Oklahoma districts responding,[7] all were regularly administering standardized achievement tests. Reading, mathematics, language arts, science, and social studies skills were being measured at a local district expense of $3-4 per student. The State Department of Education in contrast proposed three options to the legislature with an approximate cost of $4-6 per student.[8] The three options include a norm-referenced testing program, a random sample norm-referenced test, or a statewide criterion-referenced test. The state legislature chose to mandate a variation of the proposed statewide norm-referenced test for grades 3, 7, and 10. The lowest bid for the program was $125,000 or less than one dollar per student tested.

The fact that over two-thirds of the school districts were already using assessment programs and the short lead time with which the program was to be implemented indicate that information may be gathered for district and state comparisons. A second and possibly far-reaching objective may be the attempt to informally implement a statewide curriculum based on the state department "suggested" guide. The RFP mentions that the tests will be based on this recommended state curriculum.

Edward Roeber received financial data on the costs of large-scale assessment programs from 27 states.[9] The costs ranged from a few hundred dollars to several million dollars. As an example, expenditures for six states with prominent programs are shown for 1984 in the following table.

Table 1: State Costs

Florida	$4,060,965
Kansas	250,000
Michigan	1,100,000
New York	1,700,000
Texas	4,400,000
Washington	210,000

Florida's current budget for these programs is now over $6 million. The total includes tests for college students as well. Elementary and secondary education receive $2.5 million for assessment programs. The amount appears high unless compared with the total expended on education. The funds used for assessment are a small percentage of the $2.5 billion spent for education in Florida. The real cost for this program is not the testing portion but the remediation required by law. Washington spends only $1 per student in their testing program but then spends an additional $285 per student identified for remediation. Florida spends $2.5 million for state-sponsored testing but then adds an additional $33 million for state-sponsored remediation. Since remediation appears to be an important consideration for Oklahoma, an important and unanswered question is how the additional program funds will be provided for the mandated remediation program.

A final consideration may develop if the state chooses to use this assessment program as the main measure of quality for state programs. A number of issues regarding equity and adequacy could develop to further fuel current efforts to challenge the fairness of the current financial formula. The costs of these efforts could be very significant to the state.

The legislature should carefully weigh these ramifications in future actions concerning the testing program. Five important concerns should be considered: the primary purpose of the program should be improvement of instruction; the test scores are only one data set in examining accountability; the information should be used to look at the state as a whole unit and not for district comparisons; the state department and legislature should examine the success of the current program before expanding it further; and programs successful in other states may not meet Oklahoma's needs. Careful planning and use of the program should then produce several positive outcomes including allowing one more source of data to provide additional planning information to districts, requiring school districts to examine their current programs, and providing direction for support of new state programs and funding.

References

1. "Student Achievement; Test Results in the SREB States, 1985," Southern Regional Education Board, 1985, p. 2-3.

2. John M. Folks, State Superintendent, "Proposed Options, Oklahoma School Testing Program," Oklahoma State Department of Education, 1984, p. 3.

3. "Requests For Proposals For The Delivery of Products and Services For The Oklahoma School Testing Program." State Department of Education, 1985, p. 13.

4. John M. Folks, p. 3.

5. House Bill No. 1466, Section 7, 1985, p. 7.

6. House Bill No. 1466, Section 6, 1985, p. 6.

7. John M. Folks, p. 1-2.

8. John M. Folks, p. 4.

9. Edward D. Roeber. "Survey of Large-Scale Assessment Programs," Michigan Department of Education, Fall, 1984, p. 29-30.

State-Mandated Student Assessment: Some Philosophical Concerns

by

Daniel Selakovich

Nearly everyone is in favor of improving the quality of education. The disagreement comes over how that can best be accomplished. Obviously an overwhelming majority of state legislators in the United States are in agreement that state-mandated minimum competency tests of some kind are necessary and proper since more than 40 states are doing it by the beginning of 1985.

In Oklahoma the effort was launched by House Bill 1466 in July of 1985. H.B. 1466 directed the State Board of Education to implement and administer a statewide testing program designed to "evaluate the effectiveness of the public schools as shown through competence...of pupils in basic skills." Beginning in the fall of 1985, tests will be administered in grades three, seven and ten (the law exempts IEP students). The law requires testing in reading, math, language arts, communications, science and citizenship. It directs school districts to provide a program of improvement "based on the test results." The legislature appropriated $600,000 to implement the program.

State-mandated assessment is an appealing idea. If we are to provide true equality of opportunity in education, we should not only measure inputs such as per-capita spending but outputs. We need answers to these questions: Do our students have equal opportunity to achieve? What are our children learning in school? Are some learning more than others? Why?

Tests seem to be the easiest way to find answers. Unfortunately, tests won't solve the problems; they merely give us data we can use in trying to find solutions. Even this use of tests assumes that there is a consensus on what the school ought to be doing. Generally the more simplistic view one has of schooling, the more utility tests seem to have. Thus, tests make a lot of sense for those who are certain that schools should impart specific information, values and skills. This is an appealing view of schooling because there appear to be basic things every person must know if he/she is to function effectively in society. Moreover, if we stick to the really simple stuff, the things one must know are definable, quantifiable and testable. Since society is obviously less efficient and damaged by large numbers of incompetents, it follows that it is in the best interests of the individual and the society for the state to develop minimum standards and to police the schools to see that these standards are implemented. This is beginning to sound like Edward Bellamy's utopia (Looking Backward). Is it, like Bellamy's utopia, fraught with the dangers of centralism, authoritarianism and dogma? This is the broad outline of the problem. Everyone can agree that schooling is a public good, that quality is even better, and that all citizens should know the basics and be enlightened. Beyond those hopelessly vague generalizations, there isn't much agreement.

Specific Problems Inherent in State-Mandated Student Assessment

There may be so many problems with mandated student testing that it might not be worth the effort. Space will not permit a comprehensive listing, so a few illustrations will be presented. Any list would include the following: Who decides what will be tested? What effect will tests have on minorities? Social class? Teaching? the Curriculum? What uses will be made of tests? Do tests damage students? How should we test the handicapped? What are the legal problems? What will happen to local control?

Who Decides?

State-mandated student assessment assumes that somebody, somewhere knows what needs to be tested. In most states where testing has been mandated, as in Oklahoma, the mandates come from the state legislature or, through permissive legislation, the state department of education. The law can be written in some detail by the legislature or left to the discretion of the bureaucracy. The questions this raises are many: Do state legislators know what students should be studying? If not, how do they find out? Where do these laws come from? Who introduces them to make decisions about what our children should be learning? If not the legislators and their favorite lobbyists, who? Shall we leave it to the bureaucrats in the state capitol? Test experts? Teacher's unions? Education Professors? The Klan? Jerry Falwell? Who decides is vitally important and presents some very practical difficulties. In Oklahoma, the testing sections in H.B. 1466 originated with the Governor's blue ribbon committee on education. Its recommendations focused on reforms which are currently popular, such as testing, rather than current unpopular ideas such as improved programs for minorities.

The legislature of Oklahoma avoided some of the difficult questions by making the law very general. For example, they provided for tests on "communication" and "math" and "language arts" but didn't spell out what these things mean. They dumped the responsibility for deciding the specifics on the State Board, which probably means that the state education bureaucracy will make the specific decisions. This will leave a lot of unanswered questions. Where the power to decide what the content of the tests will be resides with the state bureaucracy, there doesn't seem to be any practical democratic way to involve large numbers of parents and taxpayers in the decisions. This is unfortunate in view of the tremendous impact these tests can have on what is taught and how it is taught.

How Will State Assessment Affect Race Relations?

Inkeles defined competence as "the ability to attain and perform values and rules" in society.[1] This seems to be the way the legislature is defining competence in Oklahoma. What is valued, by definition, is what the tests measure. Whoever decides, we can be sure that they will measure mainstream skills and values. This presents a problem. Researchers have demonstrated over and over that children who know only Black street culture are disadvantaged in the mainstream culture of the school. State tests can do very little to correct this problem. They can even exacerbate the problem by further convincing certain groups and individuals that they are inferior. Or worse, they can confirm the suspicions of prejudiced teachers.

There is plenty of evidence from state assessment programs which indicate that race is identifiable in the outcomes. One of the first such programs in Florida for example, reported in 1978 that in the first round of testing of 120,000 high school juniors only 23 percent of Black students passed the math part of the test compared with 76 percent of whites.[2]

There doesn't seem to be agreement in the Black community on student testing. Some feel that it can keep teachers and schools accountable, others argue that tests will be used in racist ways. Unfortunately, there is little in our past record to indicate that testing greatly improves the education of Blacks.[3] In data provided by Haney and Kinyanjui and others, it is clear that Blacks do not do as well as whites on state-mandated tests. If this were due solely to the neglect of the educational needs of Black children, then perhaps the school could identify the problems and proceed to take remedial action. This obviously is the intent of H.B. 1466 in Oklahoma. Unfortunately, this is too simplistic. Black children do not score as well as whites for many reasons--all related to racism in one way or another. Blacks tend to be poorer, come from single parent homes, culture and language is often different, there are higher levels of unemployment among Blacks, childhood enrichment opportunities are not as great and so on. None of these problems are within the control of the school. The school can only provide limited remediation for general social neglect.

However, if we use the results of achievement tests to help those who fail to gain competency in the areas tested, the results could be quite good for minorities. If, on the other hand, we neglect them or merely coach the failures to do better on the second try, or make them repeat a grade, we have hurt more than we have helped. Unfortunately, there is little in the history of testing to indicate that we will use the tests to help students. Part of this is simple racism but a good deal of it has to do with resources. Once the tests identify the students needing remediation, what then? Will the classroom space, special teachers, materials and other resources be there to do the job? History tells us no.

Oklahoma' H.B. 1466 requires remedial treatment. If taken seriously by school districts this could have ironic results. The very people most supportive of school testing--middle-class, academically-oriented educational conservatives--stand to lose the most. We know in advance those needing remediation will be overwhelmingly poor and minority children because they are the ones schools have neglected. In order to implement the language of H.B. 1466 to provide remediation, we will have to significantly increase school budgets or take resources from those children who do well on tests to provide extra help for those who do poorly. Neither of these ideas is going to be very appealing to middle-class, educationally conservative parents.

The Misuses of Tests

There is nothing in our past to indicate we have used tests to help minority children. Indeed, as hundreds of court cases illustrate, (Larry P. v. Riles is the most notable example) the opposite is true.

The National Coalition of Advocates for Students, in their report, Barriers to Excellence: Our Children at Risk (1985), found serious

misuses of minimum state competency tests. The Coalition charged that assessment of competence is political. In their words: "Elected officials and educational administrators increasingly use the scores from such tests in comparative ways." The implication is that this is a misuse of tests. According to the Coalition report: "Tests are not used to diagnose individual strengths and weaknesses, but to place children on a bell curve."[4] That is, to rank children and the ones who wind up on the bottom are always low-income, minority and handicapped children.

If low scores on tests automatically triggered remedial help, then one might be able to justify them, but this doesn't seem to be happening. What is happening is that the tests have tended to label children and restrict both children and teachers from being different and creative. The following statements from testimony by the Coalition illustrate this:

From a New York principal:

> By defining normalcy as a band on the normal curve we have seriously altered our view of children. Despite the lip service we pay to the myriad ways in which individuals differ, and claim to celebrate this variety, our practices speak otherwise. In fact it is performance on these tests--with their narrow and rigid definition of when children should be able to exhibit their knowledge--that determines whether we see children as "okay" or not. In the process we damage all children--we devalue the variety of strength they bring with them to school.[5] All differences become handicaps.

Or this from a Boston teacher:

> I like to figure out what is going on with a kid and possibly piece it back together again, but I'm bound to terms of certain objectives that the kids have to meet and I have to see that they meet those specific objectives. If they don't meet them, it falls on my shoulders. These objectives are judged by the school committees and by tests. Right now I'm really putting up a fight against these standardized tests...I dislike that type of regimented testing. I think it's very cruel...I'd rather have a curriculum that is based on the child's need....[6]

Perhaps we will be able to avoid these problems in Oklahoma.

The Effect on Teaching and the Curriculum

Of course, any teacher who is not totally stupid or suicidal or both is going to be affected by mandated assessment. It will influence what they teach and the way they teach it. We all want to look good. Since every state assessment effort has to be objective in nature and will measure those minimum competencies which are easily measured, preparing for tests tends to become rote memorization. Teachers who want to look good on state assessments will soon discover that creative teaching, innovation in the classroom and individualization of instruction will not do. The effect on methodology could be devastating. The effect on the curriculum would be to fix existing content on the schools. Tests must be designed based on what is currently being taught. State-mandated assessment tends

to fix the content the tests measure more or less permanently on the system. Change in the curriculum is difficult as it is--state-mandated student assessment can make it even more difficult. Testing can have a very negative effect in certain areas of curriculum. In the social sciences and humanities, for example, real meaning comes not from memorization of facts in the separate disciplines which make up these areas, but from the demanding intellectual task of seeing relationships which were not previously perceived. State-mandated assessment clearly discourages this approach because the real knowledge that grows out of it cannot be tested a-priori. It is interesting to note that in Oklahoma, HB 1466 is more specific about what is meant by "the principles of citizenship" than it is about the other "basics."

The Effect on Class

A major claim made for schooling in America since the beginning of the compulsory school movement was that schooling provides the best means of upward mobility. Although this idea has been seriously attacked, some still believe it. An aspect of the American school that has made it less an instrument for the reproduction of existing social class than are schools in other countries is the variety which exists in American schooling. This point was made by David Riesman when he observed that "Schools do least damage when there are many competing criteria for success--for example, the brains, the athletes, the jazz fans, the girl fans, the literary set, and so on...where there are many competing avenues to recognition no single one is apt to destroy the possibilities for differentiation by subjecting everyone to a unidimensional definition of what matters in life."[7]

This may be one of the hidden dangers in state-mandated assessment. The tests take on a life of their own. They become the major hurdle to be cleared by both teachers and students. Since we already know that there is a close relationship between social class and school achievement, we can know in advance that the poor will have the lowest scores. The tests could then become just one more device to extend the damaging impact of social class conflict which already exists in most schools.

Damaging Students

It would be difficult to imagine an assessment program that would not result, to some extent, in labeling. The schools already label kids and put so much emphasis on test performance that children classify and group each other.

The impact this has on self-image, motivation and ultimately success in school can only be guessed at, but certainly deserves serious consideration.

Although damage done bears heavily on the children of the poor, others can be hurt as well. Reisman put it this way: "Students can be damaged almost as much by success in school as by failure, for the one who is a success may be overpersuaded of his abilities along certain limited lines and underacquainted with many aspects of himself and his setting."[8] God save us from more smartass kids and their parents who think their kid's a genius.

The major damage from state-mandated assessment, however, would be indirect. As assessment influences what is being taught and the manner in which it is taught, one can expect an increasing meaninglessness of the school experience for many children. This has always been a problem--probably more in countries where testing was frequent and state-directed than in the United States where students' immediate needs and concerns have had at least a minor influence on curriculum and methods. Tolstoy noted this problem in European schools where assessments were widely used more than a hundred years ago. "In Germany," said Tolstoy, "nine-tenths of the school population take away from school a mechanical knowledge of reading and writing, and such a strong loathing for the paths of science traversed by them that they never again take a book into their hands." These were the results of emphasis on drills and testing on basic skills according to Tolstoy: "Not only does such a school breed loathing for education, but in these six years it inculcates upon these pupils hypocrisy and deceit,...and that condition of incoherence and confusion of ideas, which is called the rudiments of education."[9]

The Handicapped

Minimum state assessment faces perhaps its most serious obstacle in attempting to determine what the policies should be regarding handicapped students. Recent laws and interpretations of the U.S. Constitution minimally guarantee the handicapped the right to nondiscriminatory treatment, an "appropriate" education and substantive as well as procedural due process. In light of this, the problems with state-mandated testing are many. Certainly, some parent is going to demand that his child not be excluded from state tests, as provided for IEP students in HB 1466, on the grounds that he/she is being neglected and discriminated against.

Unfortunately, when minimum testing standards are applied to handicapped students there are many unresolved issues. According to Martha McCarthy the major unresolved issues are test validity, individualized versus standardized diploma criteria, and identification of minimum competencies. McCarthy suggested that minimum competency mandates "may ultimately force states to be more precise in articulating objectives for public schools and in defining the purpose of a high school diploma."[10]

Unfortunately, if this problem were miraculously solved it might create even a larger one. As Haney and Madaus point out:

> If minimum competency testing programs are instituted on the assumption that they assess skills necessary later in life but in practice focus on such school skills as reading and math, an interesting legal issue arises. A student whose diploma had been withheld on the basis on reading and math scores might well sue the local or state education agency on the grounds that reading and math scores have weak validity in predicting later life success....[11]

They go on to suggest that "there is little scientific basis to support the idea that we know what the important skills are...."

Legal Problems and the Loss of Local Control

The major court case on state-mandated assessment is Debra P. v. Turlington,[12] a case challenging the Florida minimum competency test. The state of Florida required students to pass a functional literacy test before being granted a diploma. Debra failed the test and was not graduated. Her lawyers argued that this was a violation of her equal protection, due process and Title VI rights, among other things. The district court agreed, saying that she failed because of past discrimination over which she had no control. The district court permitted the use of the test for remedial purposes, however. The appeals court was interested in the test itself. It wanted to know if the test was "a fair test of what was taught in Florida's classrooms." The state was able to provide proof that it was and the appeals court affirmed the state's right to deny diplomas based on the test. The Debra P. case provided other state legislatures an idea of what they must face if they decide to take this approach to minimum competency. In Florida, in order to establish the validity of the test to the court's satisfaction, the state had to hire a private consulting firm to conduct a comprehensive study which included student and teacher surveys, reports by all school districts in Florida and verification of each district's report by a team of auditors. Given the current environment in education and the availability and eagerness of lawyers, it is doubtful that any comprehensive state-mandated assessment program will go unchallenged in the courts. One begins to wonder if state assessments are really this necessary.

Also, in case you hadn't noticed, local decision-making in education has been greatly diminished. The courts, professional organizations, and regional, state and national bureaucracies have taken over. In a comprehensive study of state control of education, Frederick Wirt concluded that most of the important functions of local schools are controlled by the state. The major gatekeeping functions of accreditation, certification and attendance are completely controlled by the state. Wirt found high levels of state control on curriculum, textbooks, finances, evaluation, teacher employment and in many other areas. On Wirt's scale of 0 to 6, the median score for state control of all functions is 4.64.[13]

State-mandated assessment, where it exists, further centralizes state control over local schools. School administrators know what this means even if many local patrons don't. Its meaning was clearly stated by two school superintendents quoted in the journal, The Public Interest: "Every time there's a recognition of a problem, the legislature mandates a course," said Clair Burns, the Norton, Ohio school superintendent. "If the Legislature determines levels of competency, it won't matter too much what the communities want" adds Cuyahoga Falls superintendent Harold Wilson.[14] One could argue, I suppose, that local control has already failed. Basically, what state-mandated minimum assessment programs are saying is: Local people don't care and local school boards are incompetent.

Conclusion

Wouldn't it be nice if we could significantly improve the education of our children, especially the poor and minorities who are currently being cheated, merely by requiring the schools to test them? Unfortunately, it is not that simple. If we are really serious about improving the quality

of education, we will have to do some things which are not politically popular, such as vastly increased funding and providing programs for the poor and programs which attack racism. For teachers and some professors, the frustrating aspect of all this is that we know how to raise reading levels and improve learning in basic skills and content areas. It can be done with remedial programs, smaller classes, more teachers and better materials.

Passing another law might be politically desirable but I'm not sure it does much for education. Indeed, when one thinks seriously about the extensive nature of state control which can result from state-mandated assessment of school pupils, its effects on the lives of people, its cost, the bureaucracy that will have to be created to administer it and the legal problems created, one is reminded of this statement by G. Gilmore in the Ages of American Law: "The better the society, the less law there will be. In Heaven there will be no law, and the lion will lay down with the lamb...the worse the society, the more law there will be. In Hell there will be nothing but law, and due process will be meticulously observed."

References

[1]Inkeles, Alex. "Socialization and Schools," Harvard Educational Review, Oct., 1968, p. 60.

[2]Haney, Walter and K. Kinyanjui. "Competency Testing and Equal Educational Opportunity," Readings on Equal Education, ed. M. Barrett and Charles Harrington. New York: AMS Press, Inc., 1979, p. 345.

[3]Ibid., p. 348-54.

[4]National Coalition of Advocates for Students, Barriers to Excellence: Our Children at Risk. Boston: NCAS, 1985, pp. 45-6.

[5]Ibid., p. 47.

[6]Ibid., p. 47.

[7]Riesman, David. Constraint and Variety in American Education. Lincoln, Nebraska, University of Nebraska Press, 1956, p. 121.

[8]Ibid., p. 121.

[9]Tolstoy on Education. Translated by Leo Wiener. Chicago: University of Chicago Press, 1967, p. 13.

[10]McCarthy, Martha. "The Application of Competency Testing Mandates to Handicapped Children," Harvard Educational Review, May, 1983, p. 164.

[11]Haney, Walter and G. Madaus. "Making Sense out of the Competency Testing Movement," Harvard Educational Review, May, 1968, p. 465.

[12]Debra P. V. Turlington 644 F. 2nd 397 (5th Cir., 1981).

[13]Wirt, Frederick M. "What State Laws Say About Local Control," Phi Delta Kappan (April, 1978), p. 518.

[14]Doyle, Denis P. and Chester Finn. "American Schools and the Future of Local Control," The Public Interest, Fall, 1984, p. 85.

Draining the Swamp: Philosophical Implications of State-Mandated Competency-Based Testing

by

Kay Sather Bull

The ecological or hermeneutic history of a movement determines how it is perceived and by whom (Bruner, 1985). The title of this talk refers to this history by way of allegory; many years ago a story circulated among educators and materials developers about a project in Florida which was designed to drain a swamp. It was a long story and I will not relate it all here, but its essence, as described by the narrator, related to his ability to concentrate on the goal of the project, draining the swamp, when he was up to his hippockets in alligators.

At this conference you will meet a number of alligators who will nibble around your beliefs about Competency-Based Evaluation (CBE) or Testing (CBT) from political, economic, legal and philosophical viewpoints. This presentation then is the philosophical alligator.

CBT as a Viable Option: In Which Context?

We must start by examining the goal of state-mandated CBT's which is to adequately assess children according to or in reference to, some agreed upon standard. This goal initially makes sense from a psychometric and a philosophical standpoint if the majority of the constituency has the same definition of competency; if all children are exposed to the same curricula as the basis for testing; if the students from different districts are perceived to be a relatively homogeneous group; and if instruction is carried out in approximately the same way for all. (See Mitchell & Spady, 1978, for system details.) Generally, the above conditions are not met (Novick, 1981) and because of this, and because of philosophical differences, most educators oppose minimum competency testing (Lazarus, 1981, Glass, 1977).

Initially we must differentiate between philosophical systems related to individuals and those related to groups. The rationale for this distinction is simple; from the standpoint of the test constructor learning is an individual rather than a group activity. All assessment of the CBT variety is related to individual performance. If learning, and concomitantly assessment, were an activity which was done by a group then it would be the group rather than the individual which was assessed. All CBT assessment with which I am familiar is individual.

A variety of societal or group philosophies are defined in Table 1. Although those who preach the need for CBT may recommend one or more of these philosophies it can be seen from the definitions that none are truly appropriate for CBT. Nor would a psychometrician accept a group philosophy for an individual activity. The challenge from a philosophical view to CBE and CBT begins by questioning the appropriateness of the goal from the standpoint of various philosophies.

Table 1

Egalitarian: To each person an equal amount of the educational resources.

Utilitarian: To each person to maximize societal development either average or aggregate utility.

Libertarian: To each person according to societal contribution or to them who has.

Meritocracy: To each person according to his effort or merit.

Socialist: To those according to their needs; from those according to their abilities.

* From Bull (1985)

Arguments for a Philosophical Base for Assessment

The goal of CBE and CBT is to provide assessment of what the schools should teach. From a philosophical standpoint this goal is open to two sets of challenges: (1) should CBT be used, an issue related to choice of philosophy, and (2) what should be taught, an issue determined by choice of philosophy.

First, I would like to establish that having a philosophy is important and then I will show the outcome of adapting a philosophy for advocates of a CBT program.

Combs (1965) asserts that what we teach, the way in which we teach it and the way which we assess what is taught depends on our beliefs about what people are like. In the same manner, the goals we seek, the kinds of things we do with others, the judgments we make, and even the kinds of experiments we are willing to try are determined by our beliefs regarding people, society, and the world. When we organize some subset of our beliefs into a set of warrantable assertions, for which supporting evidence (both rational and empirical) can be provided, we have formed a theory (Morris & Pai, 1976). For example, we may have a theory about the role of the school in transmitting culture, a theory about the nature of the learning process, a theory of motivation, and a theory regarding the way or ways in which we evaluate learning. Our theories (as organized sets of beliefs open to critical inspection) and our practices in educational settings interact with each other. That is, we check and criticize our practices on the basis of theory. This latter process, the evaluation of practice from the perspective of theory, is an important one that is often undervalued by contemporary educators and psychometricians. They often proceed as if they can evaluate the validity of an educational or psychometric practice independently from the theoretical framework in which it is embedded.

Educational theories should serve as a guide to practical action, especially when a problem arises for which no immediate solution is available. A common difficulty that occurs in this context, however, is the reliance on two or more theories that are incompatible with one another. There is a tendency among educators to be eclectic; that is, to pick and choose

various theories and procedures without sufficient consideration as to whether these theories and procedures are mutually compatible. Bigge and Hunt (1980) describe this phenomenon as follows:

> Eclectics who are not extremely astute in identifying their deepest assumptions may very easily--and with good intentions--select from differing schools of thought ideas that on superficial examination seem very attractive. But suppose, when examined in the light of their primary assumptions, these ideas are contradictory. This is a risk that eclectics take, and they often fall into the trap of building a new point of view out of incompatible elements. What emerges is a position that may at first glance look logical, but on more careful scrutiny is a hodgepodge of contradictory and mutually exclusive components. (p. 6)

Such eclecticism, when it influences the practice of classroom teachers, may result in one or more of the following practices: Teachers may use one method while verbalizing another. The teacher's goals may be poorly conceived, resulting in the assignment of aimless busywork for students. Teachers may speak of the importance of self-fulfillment and then force their students to adhere to a tightly prescribed curriculum (Shermis, 1967).

From a psychometric standpoint an eclectic may prescribe tests which have single right answers; e.g., multiple choice, true/false, matching; in situations which, from a philosophical standpoint, require divergent production and synthesis; e.g., reflective, problem-based, essay examinations. Psychometricians may also prescribe assessment or evaluation in situations where philosophically no assessment is tenable.

There are many examples of eclecticism which can be found in the writings of test constructors and consumers. One such example comes from the literature which advocates the use of multiple choice (standard level) tests as a measure of the capability to solve problems involved in living in the real world; e.g., tests of basic life skills. The assumption in basic skills curricula is that students will be capable of functioning as problem-solvers in, for example, job-related experiences, when they can pass a multiple-choice examination when, in fact, hands-on performance would generally be considered a better indicator. Another inappropriate usage occurs when students are given an essay examination which is timed and on which they are graded or evaluated on the basis of facts included within the response. Obviously the assessment of facts is best left to multiple-choice or other forms of recognition response items.

In order to avoid the difficulties engendered by eclecticism, numerous philosophers and educational theorists have attempted to expand and synthesize their theories into comprehensive and coherent systems of educational thought. These systems serve to organize specific beliefs and propositions into an organized whole. This meaning of the word "system" is consistent with that found in The American College Dictionary (Barhart, 1961): "An ordered and comprehensive assemblage of fact, principles, doctrines, or the like in a particular field of knowledge or thought...."

As we have tried to point out, the field of psychometrics is not immune from difficulties encountered by taking an eclectic stance to educational theory. The kind of testing used, as we shall see, is determined by the philosophical system. The kind of assessment used must be matched to the curricula and that, in turn, is prescribed by the philosophical system.

We now turn to a necessary brief discussion of four such systems and the implications of creating a testing or assessment system within each of the systems.

Psychometrics as Conceived Within Four Systems of Educational Thought

Psychometric assessment, if it is to be practical and excepted by knowledgeable persons, should be embedded in a system of educational thought. Of the many existing systems, (see, for example, Bruner, 1985) four have been chosen for discussion. It is believed that these four, and variations of them, represent a significant cross-section of existing positions. The major viewpoints that will be considered are mental discipline theory, behaviorism, cognitive-field theory, and humanism. The discussion of these viewpoints relies heavily on their presentation by Morris Bigge (Bigge, 1982; Bigge & Hunt, 1980). (See Appendix A for an expanded presentation of each of these educational viewpoints, or systems.)

Mental discipline theory served as a foundation for the prominent approach to education in the United States and Europe during previous centuries and its influence continues to be felt in contemporary practice. Discipline theory assumes that the moral nature of persons is inherently bad and that dull and boring activities are necessary in order to develop "faculties" and to suppress bad tendencies in the student. The discipline theory approach calls for large-group instruction and "more of the same" in terms of content. Learning is based on repetitive drill and rote memorization; learning is assessed by determining if the student has retained factual material. Assessment is accomplished by having the students give back the information; e.g., factual essay, short answer, true/false or completion items.

A second major viewpoint is that of behaviorism. Within this perspective, the focus is on arranging environmental events to produce a specific change in behavior. Behaviorism takes a neutral stance with respect to the moral nature of persons. It views people as passive and reactive organisms and calls for the shaping of desired behavior using reinforcement and the presentation of information in small steps. From a behavioral perspective, assessment requires a sampling of desired responses. For example, if the objective to be tested was to balance a checkbook, the student would be given problems from a sample checkbook, not items to assess basic addition and subtraction in other contexts. Appropriate test items for this philosophy are determined by the kinds of responses required by the S-R situations previously presented. Only teacher determined response items are appropriate.

The third perspective, cognitive-field theory, views the moral nature of persons as neutral (neither good nor bad). Unlike behaviorism, however, cognitive-field theory views the actional nature of persons as interactive and purposive. Learning is viewed as a negotiated process between

the teacher and the student. Critical to cognitive-field theory is the perspective and belief system of the student. Assessment in this system is usually individualized or personalized. The student is presented with real, unanswered, problems in his/her area of interest or study. These problems involve both generalization, skill application and creativity. The form of the test item is the reflective or problem-centered essay.

The fourth system, humanism, assumes that the learner is the source of all learning, that truth is immanent, and that external agencies have no business doing more than removing roadblocks from the self-development of the learner. Because each learner is unique there is no assessment other than self-assessment. There is not nor can there be a valid, systematic, testing program. Each learner defines his or her strengths and weaknesses.

It is important to note that within the educational theories we have discussed above, there is little foundation for CBT as we have come to know it. The closest philosophy to CBT or CBE is behaviorism yet this philosophy has never been adapted within the public schools. If it were, education would accept students at the level at which they were functioning and would provide an appropriate menu with which to proceed from that point. Considerations about length of training, disabilities, remediation, etc. would be moot in a behavioral education setting as would any such requirement for a statewide CBT, at least from an educational viewpoint. Politics, however, makes a strange bedfellow and behavioral philosophies, because they bow to power, will, like the lamb, lay down.

A Nonphilosophy Philosophy

One further philosophical position merits brief examination, that of the reconceptualists. Apple (1979) states that society provides different educational experiences for students of different classes. Those with lower ability or lower class are taught punctuality, neatness and habit formation while those with higher ability/class have their curriculum organized around flexibility, choice and inquiry. The reconceptualists argue that education is designed to replicate the existing social class structure and argue, therefore, that to change this policy we should provide the most for those who are the least prepared (cf, Rawls 1971).

CBT, according to the reconceptualists, is a device used to exclude those from the lower classes from educational advancement. Therefore, it should not be used. Schools, according to Tunnell (1978), should cease to track and should not use tracking mechanisms; e.g., CBT, because they (help to) reproduce economic inequality by propagating ideological beliefs that are supportive of capitalism. Inherent in this belief is the need for a restructuring of society away from present forms toward a more equitable (socialistic) form of political and economic process with concernitant persons of educational policy.

Summary

Two quotations from different authors seem to adequately summarize the philosophical view toward CBT. The first from Lazarus (1981):

> Minimum competency enacts into law...the philosophy that a certain kind and amount of education is correct for everyone. The

> legislation stems explicitly from the notion that no person can succeed...without that minimum. It is quite a piece of effrontery, if one stops to think about it, to maintain these skills, and no others, are what a person absolutely requires to make his way in the world. (p. 189)

The second from Mitchell & Spady (1978):

> If the pressures for school reform and redirection are to be creative rather than merely disruptive, it is vitally important that research and policy analysis focus on the relationship between the increasingly explicit political demands for educational outcomes and the specific organizational changes required if the schools are to pursue those outcomes. Without a clear understanding of how organizational mechanisms (philosophies) facilitate or inhibit desired student outcomes, pressure for improved performance can only lead to frustration and disappointment. (p. 17)

A quick tour of the philosophical thickets has indicated that none of the extant philosophies, either societal or individual, is appropriate for CBT with the possible exception of behaviorism. Behaviorism, however, because of its underlying assumptions, violates the very essence of a democratic society (Millman, 1980) and, if used, would find no educational rationale, albeit possible political rationales, for the use of CBT's.

From the philosophical view, then, this particular swamp should not be drained; instead add alligators and CBT may be consumed along with other things which have been used historically to fill the swamp.

References

Barnhart, C. L. (Ed.). (1961). The American College Dictionary. New York: Random House.

Bigge, M. L. (1982). Learning Theory For Teachers (4th Ed.). New York: Harper & Row.

Bigge, M. L., & Hunt, M. P. (1980). Psychological Foundations of Education. (3rd Ed.). New York: Harper & Row.

Bruner, J. (1985). Models of the Learner. Educational Researcher, 14 (6), 5-8.

Bull, K. S. (1985). Equity & Justice in Gifted Education: Implications for Curricular Philosophy. Paper presented at the annual meeting of the American Educational Research Association. Chicago.

Combs, A. W. (1965). The Professional Education of Teachers. Boston: Allyn & Bacon.

Glass, G. V. (1977). Standards & Criteria: An Occasional Paper. Kalamazoo: Center for Evaluation, Western Michigan University.

Lazarus, M. (1981). Goodbye to Excellence: A Critical Look at Minimum Competency Testing. National Association of Elementary School Principals, Boulder, CO: Westview Press.

Millman, J. (1980). Implications of Minimum Competency Testing Programs for Achievement and Achievement Testing. (pp. 31-38). In W. B. Schrader (Ed.). Measuring Achievement: Progress Over a Decade. New Directions for Testing and Measurement, 5, 1980. San Francisco: Jossey-Bass.

Mitchell, D. E., & Spady, W. G. (1978). Organizational Contexts for Implementing Outcome-Based Education. Educational Researcher, 7 (7), 9-17.

Morris, V. C., & Pai, Y. (1976). Philosophy and the American School (2nd Ed.). Boston: Houghton Mifflin.

Novick, M. R. (1981). Federal Guidelines & Professional Standards. American Psychologist, 36, 1035-1046.

Rawls, J. (1971). A Theory of Justice. Cambridge, MA: Harvard University Press.

Shermis, S. S. (1967). Philosophical Foundation of Education. New York: Random House.

Tunnell, D. R. (1978). An Analysis of Bowles' & Gintis' Thesis that Schools Reproduce Economic Opportunity. Educational Theory, 334-342.

Minimum Competency Testing: Some Practical Concerns

by

Janie L. Hall

The reasons for implementing minimum competency programs are laudable and the concept seems straight forward; the public wants a way to hold educators accountable and educators want to improve their delivery of education to students.

But this apparent simplicity is misleading. After more than a decade of controversy, the notion of minimum competency testing (MCT) is still full of contradictions (Pipho, 1978) with neither the "pros" nor the "cons" making any progress in retiring the opposition (Stronge, 1984).

And while the battle over the relative merits of MCT goes on, states are continuing to adopt programs. The number of states that can be identified as having a MCT program seems to depend on which report one reads. The Education Daily reported in January, 1984, that 45 states had some sort of basic skills assessment program. Two additional states required local testing and one state facilitated voluntary testing. Of the states with mandated testing programs, 26 (58%) used the results as a competency measure ("state use," 1984). Pipho, also in 1984, identified 40 states as having minimum competency standards. Since that time, one state has considered dropping its competency requirements (T. Endo, personal communication, August 26, 1985) and Oklahoma has adopted a state assessment program. So it seems rather difficult to "tell the players without a program."

One difficulty that we have is a lack of consistent terminology. We tend to use "state assessment program" and "minimum competency testing" interchangeably. In many cases they are synonymous but not always.

A review of the literature (Perkins, 1982) revealed that the term "state assessment program" is generally defined as serving the purpose of monitoring student achievement. A student does not have to meet a specific standard. That is, students neither pass nor fail the tests. Results are generally reported by group rather than by individuals. The purpose of such a program is monitoring.

MCT goes considerably further. It involves testing for mastery of certain skills and students are expected to achieve a certain level of performance. The failure to achieve competency carries specific consequences such as in-grade retention or withholding of the diploma (Perkins, 1982). Competency tests can also provide a basis for remediation of academic deficiencies (Cohen & Haney, 1980).

I point out the distinction between the two concepts because Oklahoma has just adopted a state assessment program. It requires the administration of a norm-referenced achievement test annually, beginning Spring, 1986. A criterion-referenced test is to be implemented by 1989 (Educational

Improvement Act, 1985). Uses of the data have not been identified (at least publicly). So my argument will be that we should avoid moving from a state assessment program into a minimum competency program and thus avoid some of the accompanying negative effects that other states have experienced.

To describe what I believe we are going to have to do with our assessment program, I parody the words of a song from the 1940s, "You got to accentuate the positive and eliminate the negative" (Mercer & Arelen, 1944).

Proper Use of Tests

The National Council on Measurement in Education (1980, p. 4) has stated that "Tests can provide assessments which in many instances are more objective and more equitable than other forms of evaluation. The extent to which tests can contribute to sound educational decisions depends on the intrinsic quality of the tests and on...use of the scores."

Regarding use, most test manuals contain a statement such as the following disclaimer which appears in the California Achievement Tests, Test Coordinator's Manual, "...tests cannot measure total achievement in any school or grade, or in any discipline within a grade. For this reason, the results of these tests should not be used as the sole basis for evaluation.... The proper use of results demands recognition of the limitations and restrictions of the tests" (CBT/McGraw-Hill, 1977, p. 42).

There are many legal, philosophical, and theoretical concerns related to MCT that are being addressed by others in this conference. Therefore, the emphasis of this paper will be on what I consider the misuse of tests. My major point will be that the use of the test scores alone to determine competency constitutes a misuse.[1] I will point out the positive effects that MCT is intended to have, the negative effects of using scores alone, and then make some recommendations.

Accentuate the Positive

Certainly it is a positive thing for a State to know how its students are achieving. And educators do need to renew public confidence by demonstrating a willingness to set standards and to be accountable (J. M. Maddox, personal communication, June, 1985; Perkins, 1982).

MCT can facilitate a careful examination of the curriculum (Kohlfeld, 1984) and provide a framework for specifying what skills should be learned (Stronge, 1984). It can assure that material is covered (Darling-Hammond & Wise, 1985) and stimulate establishment of remedial programs for students who show deficiencies.

Finally, MCT can provide motivation for teachers and students, and it can give a sense of common purpose (Lambert, 1979). Some believe that the positive effects of MCT are already evident in the schools and are demonstrated by standardized test gains (Fisher, 1980).

[1] The views expressed do not necessarily represent policies of the Oklahoma City Public Schools.

Eliminate the Negative

It will quickly become apparent that the negative section of this paper is much more lengthy than the positive. While this may seem to be "accentuating the negatives," I hope that by concentrating on them in a setting such as this conference we can eliminate, or at least diminish, their effect on students.

The four areas of difficulty related to test use which I will discuss are as follows:

The inappropriateness of basing standards on test scores alone;

The added burden which use of test scores alone places on minorities and handicapped;

The undesirable impact on curriculum resulting from use of test scores alone; and

The effects of overtesting on local school districts.

Standards and Test Scores

MCT assumes that there is some standard against which a student's performance can be measured. Standard setting occurs in two ways. First, we define those skills that students who come through our schools should have. And then, we define the expected level of expertise that they should have.

Defining Skills. There is a general lack of consensus as to what life skills are important; i.e., what skills to test (Miller cited in Perkins, 1982). Some believe that schools may have been overly ambitious and attempted to identify competencies that the majority of students cannot achieve (Airasian, Pedulla, & Madaus cited in Perkins, 1982).

Conversely, decisions are being inappropriately made on the basis of objectives-referenced tests disguised as criterion-referenced tests. That is, we have made competencies trivial in order to be able to measure them. Here is what I mean.

A criterion-referenced test is used to determine a person's status with respect to a well defined field of knowledge that a student should know (Popham, 1975). Take for example, mathematics. Theoretically, there is a near-infinite number of skills in this field of knowledge; e.g., addition of four digit whole numbers or subtraction of mixed fractions. Questions can't be asked about every one of the skills because, first of all, we haven't identified them all. And besides, it would make a very long test. So we select what we consider a representative sample and infer that we are testing the whole field. If a student passes an acceptable number of items, we assume that he/she has mastered the field (Nunnally, 1978).

With an objectives-referenced test, no field of knowledge is specified and test items are not considered to be representative of any particular field. Therefore, decision makers should not make generalizations beyond

the passing or failing of the items themselves (Hambleton & Swaminathan, 1978). The difficulty is that they often do.

A different problem related to what skills we expect to be mastered is that sometimes standard setting and fairness to the student can be at cross purposes. For example, one state defined the skills that would be required for receiving a high school diploma but they began administering the test at ninth grade in order to allow time for remediating students who failed (P. Mazza, personal communication, April 21, 1985). Granted, high school students usually have acquired the basic skills by ninth or tenth grade and after that they tend to take advanced or vocational courses, depending on their track. But the public thinks of the test as a ninth-grade test and wonders why schools only expect a ninth-grade performance of their graduates (D. Covert, personal communication, September, 1985; P. Mazza, personal communication, August 21, 1985).

Defining Level of Skill. Nearly everyone agrees that we should have standards but few agree on what those standards should be. There is no natural or obviously correct passing score to separate the competent from the incompetent. As a result, a student can be competent in one state and incompetent in another (Linn, Madaus, & Pedulla, 1982). Cutting scores can vary widely depending on the procedure used to set the standard and the category of people who make the judgments. In a study of nearly 200 people, using both statistical techniques and expert ratings on the same data, group failure rates varied from 9% to 78% depending on the process used to set the cut-off score (Halpin & Halpin, 1982). Obviously, it matters a great deal to students which method is used.

A well researched review of standard-setting models (Glass, 1978) concluded that all attempts to derive criterion scores are either obviously arbitrary or derived from some arbitrary set of principles.

To highlight one of the more blatant examples of the arbitrariness of cut-off scores: In one state, it was noted that when a large percentage of students passed the initial reading and mathematics tests, there was an immediate talk of raising the passing scores so more students would fail. A state employee observed, "The MCT must be constructed so that initially enough children fail, to satisfy the people who called for the test in the first place and at the same time enough children pass so it doesn't wreck economic havoc on the schools" (National Institute of Education, 1981, p. 733).

Compounding our inability to set an absolute passing score is the matter of individual student variability. What happens to the usually good student who has a cold on test day and doesn't make the cutting score? Proponents of MCT would say that he/she just enters a remedial program then takes the test again. That is a waste of student and teacher time. Given the nature of assessment, there will always be competent students who are labeled incompetent and incompetent students who are considered competent but if we base our judgments on test scores alone, our chances of mislabeling are greatly magnified.

Special Groups and Test Scores

I would like to paraphrase a line from the play Raisin in the Sun that I think is particularly poignant and appropriate to the next negative statement about MCT: "When you start measuring somebody, measure him right. Make sure you take into account what hills and valleys he came through before he got to wherever he is" (Hansberry, 1959).

MCT assumes that all students have had the same opportunities and educational experiences. Many maintain that this is a faulty assumption and that MCT serves to compound the injury of prior discrimination, and other social and economic failures (Glass, 1978).

Minority Races. It has been demonstrated that failure rates are higher among certain minorities than among whites (Daub, 1983; Fisher, 1980; Serow, et al, 1980). Additionally, the relative percentage of minority students and white students identified as incompetent can vary with the cut-off score that is selected. For example, by raising the cut-off slightly, one can identify 3% more whites and 15% more blacks.

Linn et al (1982) point out that the testing inequities for minorities result from the combination of a rigid and arbitrary cut-off score coupled with the dependency of the score (in many cases) on a few bad items on the test. While we have made giant strides toward the elimination of biased items, we have a way to go.

Handicapped. Of equal concern for purposes of this discussion is the inclusion of handicapped students in MCT requirements. Although Oklahoma has exempted its special education students, some other states require the participation of most of its handicapped students while still others exempt most. In a study of over 1,500 students in North Carolina, it was found that handicapped students in general, and educationally mentally handicapped students in particular, scored lower on the initial examination, withdrew from school more frequently after a MCT failure, and were less likely to attain the required level of mastery prior to graduation time (Serow & O'Brien, 1983).

Perhaps a portion of the difficulty lies in the test setting itself. Traditionally, handicapped students have been tested under very different conditions from those of the group test setting that the "average student" is accustomed to; i.e., individually administered tests as opposed to group administered tests. This can be an additional disadvantage for the handicapped student who, because of an individualized educational program (IEP), may not have been exposed to the same material or in the same manner.

Dissimilar School Districts. Somewhat related to this discussion of equal opportunity are the findings of an exploratory study in Florida (Jolly, 1985). When looking at third grade scores on the state test, across seven years, it was noted that medium-sized school districts (10,000 to 49,999 students) always scored highest. Small districts always scored lowest. It was suggested that comparisons of districts of dissimilar size, using test results alone, may lead to faulty assumptions about differences in the quality of instructional programs. MCT assumes that all systems have equal resources.

Curriculum and Test Scores

Opponents of MCT argue that it leads to teaching only those skills that are measured by the test and that schools could become staffed by people who are not teachers but who are implementers of MCT's. (Crain & Watson, 1983, 1984.)

When a sample of teachers from three Mid-Atlantic states was asked about the effect of competency testing on their curriculum, sixty percent said that increased emphasis on tests to make decisions about students affected their own teaching. Ninety-five percent said that it affected other teachers. The effects they named were altered curriculum (some considered this positive, most did not), teaching students to take tests, teaching the test, having less time to teach, and feeling under pressure (Darling-Hammond & Wise, 1985).

Cramming. A carefully controlled study of the factors affecting changes in the instructional process showed state-mandated competency tests significantly changed the instructional process. The modifications were viewed as the result of curriculum changes, teacher workshops on MCT, and mandatory practice tests for students. Only the practice testing increased the measured level of basic skills by a sizeable amount. Such practice is generally viewed as cramming. A test that is susceptible to cramming may be unreliable (Walsted, 1984).

Cheating. Proponents of MCT and opponents, alike, admit that MCT does increase the incidence of cheating on the part of teachers and students (J. M. Maddox, personal communication, June, 1985; T. Endo, personal communication, August 26, 1985).

Logistics and Instructional Time

In one state where students begin attempting a test to determine eligibility for a diploma as early as grade nine, students needing to take the test a second and third time may spend as much as sixteen hours being tested. At the building level, there is a major intrusion on teachers' instructional time of two to three days, twice a year--perhaps nearly a week of instruction lost (T. Mazza, personal communication, August 21, 1985).

Closer to home, a recent article in the Daily Oklahoman (Killacky, 1985) reported that state officials have announced that the norm-referenced test called for in House Bill 1466 will be administered March 18-20, 1986, or else March 25-27, 1986, depending on spring break. Or if there is any delay, test dates could be moved to early April.

Oklahoma City administers its norm-referenced test to grades K-11 beginning in early April. We use the scores, in conjunction with other data, for a variety of individual and program decisions. We go to great lengths to get accurate data. This even includes reminding parents to send their children to school rested and fed (when possible). We are especially careful with school meals on test days, and we avoid other kinds of tests and out-of-school activities on these days. We go out of our way to get good data which is in turn used to diagnose and remediate weaknesses. So

we are more than fulfilling the apparent intent of our state assessment program.

I am concerned about what will happen to this positive testing atmosphere for students and teachers in grades three, seven, and ten who will be involved in the state test as well. If the state test is a different one, this means two major tests within a month's time. Further, in grade ten there is the additional problem of pulling students out of classes, when they are with other grades, to a central area and providing an adequate number of teachers for approximately five hours of extra testing. Besides fatigue, we again have a considerable loss of instructional time.

Finally, (and this isn't the state's fault, but it is an example of what can happen) seven of our schools are in the National Assessment of Educational Progress sample this year. Some students will take three tests this spring.

Recommendations

Please do not interpret the remarks which follow as agreement with MCT. Although I see the need for a state to know how its students are achieving, a careful review of the relevant literature has convinced me that MCT is full of dangers for individual students.

Further, a state could acquire data for monitoring its students fairly easily and relatively inexpensively by requiring all districts to administer a (any) norm-referenced test. Results could be aggregated at the state level using a common metric such as normal curve equivalents. These have been used successfully at the federal level for years.

Here, in Oklahoma, since we already have a state assessment program, my recommendation is that we not go further. Specifically, don't base decisions about promotion, graduation, merit pay for teachers (or districts) on test scores alone. Tests were not designed to be the sole criterion. Rather, establish a broad set of requirements. Grades, attendance, previous test results, and teacher checklists of student performance could be included in the decision-making model. If we were to go the MCT route, other recommendations for consideration would include the following:

> Let local districts set the minimum competency standard (Kohlfeld, 1984), with the state maintaining a monitoring function.
>
> Establish a waiver based on a previous test. For example, any student who achieves the fiftieth percentile or higher on a previous norm-referenced test in reading and mathematics would be exempt from MCT (T. Endo, personal communication, August 26, 1985).
>
> Give advance warning. For example, beginning in grade six, notify parents that the student is being passed but may have trouble with the exit test. This would benefit parents and students and free the district from possible complaints (P. Mazza, personal communication, August 21, 1985).

Don't implement too quickly. Allow at least one year to pilot the criterion-referenced test, to do extension item analysis, and to check for bias (S. Johnson-Lewis, 1983).

Make an avenue for protest by students, teachers, and parents (Madaus, 1981).

Include community representatives, such as large employers, and a large percentage of teachers in the selection of competencies considered important (Lambert, 1979). A careful study of Jaeger's model is urged (Jaeger, 1982).

This paper has dealt only with concern over the misuse of tests; i.e., using them alone. My hope is that when it is combined with papers dealing with legal, educational, philosophical, and test developmental concerns that policy makers will use the information to answer these four questions:

Is MCT technically feasible?
Will MCT lead to attainment of the desired goals?
Will other, unintended, effects occur?
Is there a better way to reach these goals?

By looking realistically at standards for our students, I believe we can accentuate the positive and eliminate the negative.

References

Cohen, D. K., & Haney, W. (1980). Minimum, competency testing, and social policy. In R. Jaeger & C. Tittle (Eds.), Minimum competency achievement testing. Berkeley, CA: McCutchan.

Crain, John A., & Watson, Hoyt F. (1983/1984). Minimum competency testing: Some arguments against. The High School Journal, December/January, 111-115.

CTB/McGraw-Hill. (1977). California achievement tests, test coordinator's handbook. Monterey, CA: CTB/McGraw-Hill.

Darling-Hammond, Linda, & Wise, Arthur E. (1985). Beyond standardization: State standards and school improvement. The Elementary School Journal, 85 (3), 315-336.

Doub, H. R. (1983). Competency tests: bane or bust? The High School Journal, April/May.

Educational Improvement Act of 1985. (1985) Oklahoma House Bill 1466. Sections 6-11.

Fisher, T. H. (1980). The Florida competency testing program. In R. Jaeger, & C. Tittle (Eds.). Minimum competency achievement testing. Berkeley, CA: McCutchan.

Glass, G. V. (1978). Standards and criteria. Journal of Educational Measurement, 15, 237-262.

Hambleton, Ronald K., & Swaminathan, Haribaran. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. Review of Educational Research, 48 (1), 1-47.

Hansberry, Lorraine. (1959). Raisin in the Sun. New York: Samuel French.

Halpen, Glennelle, & Halpen, Gerald. (1982). Setting performance standards: A documentation and discussion. Education. Winter, 118-128.

Jaeger, Richard M. (1982). An iterative structural judgmental process for establishing standards on competency tests: Theory and application. Educational Evaluation and Policy Analysis, 4 (4), 461-475.

Johnson-Lewis, Sharon. (1983). American Educational Research Association, Annual Conference. Montreal, Canada. Unpublished manuscript.

Jolly, Jean. (1985). Relationship of district size and state assessment test scores. Unpublished manuscript.

Killacky, Jim. (1985, August 30). State students face spring tests. Daily Oklahoman, p. 1.

Keefe, J. W. (1980). Differentiated diplomas: To do or not to do? NASSP Practitioner, May.

Kohfeld, Barbara Capshaw. (1984). An analysis of opinions regarding minimum competency testing. Education, 105 (1), 53-57.

Lambert, Don. (1979). Houghton Mifflin Invitational Conference, Iowa City, IA. Unpublished manuscript.

Linn, Robert L., Madaus, George F., & Pedulla, Joseph J. (1982). Minimum competency testing: Cautions on the state of the art. American Journal of Education, November.

Madaus, George F. (1985). Public policy and the testing profession--you've never had it so good? Presidential address delivered to the National Council on Measurement in Education. Chicago, IL, April 2.

Mercer, Johnny, & Arelen, Harold. (1944). You've got to accentuate the positive. New York: Edwin H. Morris & Co.

National Council on Measurement in Education. (1981). Transcript of Minimum competency testing: Clarification hearings - Washington, D.C., July 8-10, 1981. Washington, DC: Anderson Reporting Co.

Nunnally, Jim C. (1978). Psychomaetric theory (2nd ed.). New York: McGraw-Hill.

Perkins, Marcy R. (1982). Minimum competency testing: What? Why? Why not? Educational Measurement Issues and Practices, Winter, 5-9.

Pipho, C. (1978). Minimum competency testing in 1978: A look at state standards. Phi Delta Kappan, 58, 585-587.

Pipho, C. (1984). State minimum competency action. Education Commission of the States, Department of Research and Information. Washington, DC.

Popham, W. J. (1975). Educational evaluation. Englewood Cliffs, NJ: Prentice-Hall.

Serow, Robert C., Davis, James J., & Parramore, Barbara M. (1982). Performance gains in a competency test program. Educational Evaluation and Policy Analysis, 4 (4), 535-542.

Serow, Robert C., & O'Brien, Kathleen. (1983). Performance of handicapped students in a competency testing program. Journal of Special Education, 17 (3), 149-156.

State use of basic skills tests, 1983. (1984). Education Daily, January.

Stronge, James H. (1984). Minimum competency testing pros and cons: Is there really a difference? The Clearing House, 57, 217-219.

Walstad, William B. (1984). Analyzing minimum competency test performance. Journal of Educational Research, May/June, 261-265.

Competency-Based Measurement: A Review of Psychometric Concerns

by

Dianna Newman

Among the many changes of educational measurement in the 1970's and 1980's, one having a major impact has been an increased focus on the development and use of competency-based tests. Evidence of this increased attention can be found in both the theoretical and applied fields of educational measurement. Since Popham & Husek (1969) began the theoretical dialogues on these tests, measurement experts at both the academic and classroom levels have proposed solutions, developed theories, and fought professional battles as to the future of this mode of measurement. For those professionals charged with the responsibility of maintaining the psychometric integrity of education the discussions have proven both challenging and painful. Theorists are still attempting to establish the merits of competency-based measurement and researching methods of finding its psychometric properties. The lay public, however, is not willing to wait for this development and is urging, and/or mandating, the use of competency-based measurements. This "cart before the horse" phenomena of legislated application before development has forced many teachers, evaluators and administrators to go for their own knowledge of measurement and to use tests which have little or no supporting psychometric properties (Hambleton, 1980).

The debate over the usefulness of these instruments and the resulting decisions are best summarized by Perkins (1982). These issues range from cost effectiveness to public awareness to use. As noted by Perkins, however, these points reflect the issues raised by those not responsible for the psychometric integrity of educational measurements and the decisions based on those scores. The purpose of this paper is to present some of the psychometric issues faced by professionals who are working with competency-based measurement. Among the many psychometric properties, impinging on the successful utilization of competency tests, four appear to be dominating the current round of psychometric debates. Three of the four, representing an attempt to establish traditional psychometric properties, are: test length, reliability, and validity. The fourth, establishing cutpoints or standards, is a direct outcome of one of the uses of competency tests. For each issue, this paper will present an overview of the problem, a summary of the current debate and references where alternate technical solutions may be found.

Definitions of Competency-Based Tests

Prior to discussing the properties of competency-based tests, a clarification of the many definitions of the test is needed. Popham (1978) defined these tests by indicating their use: "Criterion referenced test is used to ascertain an individual's status with respect to a well defined behavioral domain" (pg. 93).

Berk (1980) indicates that, while accurate in some settings, Popham's definition really defines "domain-referenced" tests. This subset of tests has as its major concern the establishment and use of stable, precise specifications, grounded in the domain, which can be used to maximize the interpretability of an individual's score in the domain. That is, individuals' scores can be used to determine where the student fits in the range of knowledge for that domain. An alternative type of test is the mastery test. These tests are used to classify students as masters or non-masters of an objective. For mastery tests, the setting of absolute performance standards and the estimate of classification errors are necessary. Berk indicates that each of these subsets will require the presence of different psychometric properties. Traub and Rowly (1980) reach a similar conclusion when they present two questions (See Figure 1) which test developers must answer prior to determining psychometric properties.

1. Is the variable to be derived from the test continuous or dichotomous?

2. Does the psychometric property support a test score that will be used for a decision or a score that will be used to describe performance?

   (Domain-referenced or mastery?)

In answering the first question the test developer or test user must examine the relationship among the items. If the variable or final score is dichotomous, the test is measuring the student's ability to perform all or none of the items representing the domain. If the ability is not there, the student cannot perform any of the items except at chance level. This means the true score is either 0% or 100%. For continuous scores, however, each item taps certain capabilities which are required while overlapping with skills from other items. The items are not perfectly correlated within the domain. The true score can in fact range from 0% to 100%. In answering the second question, the test developer or user is concerned with the use of the score. If the use is to compare a student's acquisitions of the domain with the domain, the test is being used for descriptive purposes and is domain-referenced. If, on the other hand, the use of the test score gives beyond this description and indicates that the acquisition "is enough," or that mastery has occurred, a second purpose of the test is involved. Answers to both of the questions will affect the method of proving presence of psychometric properties.

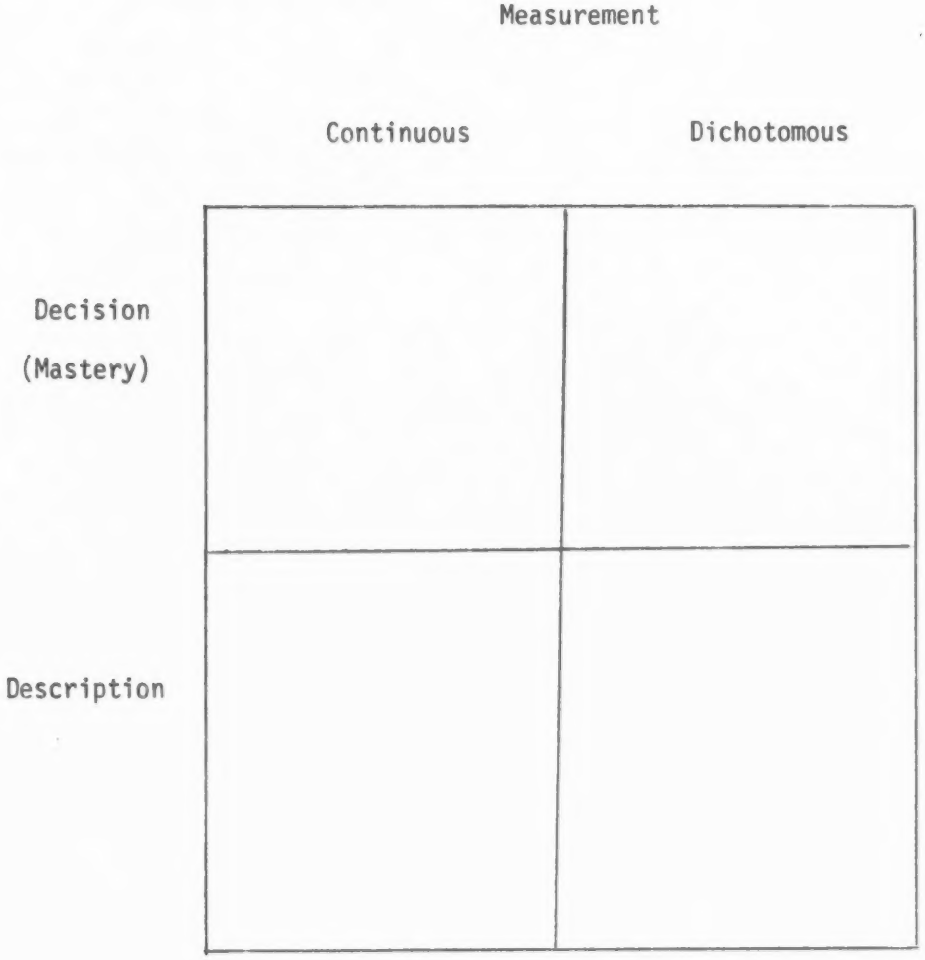
Test Length

One of the first properties to be determined by any test developer is "how many items"? There are few cognitive areas that can be so concretely defined as to allow the development of the universal population of items. Test developers need some way of establishing a recommended number of items per objective that should be written. Each of the situations found in Figure 1 may yield a different answer. In addition, other questions specific to the use of the measure may impinge on these answers. For example:

1. What is the minimum level required; i.e., what is the level of proficiency (continuous variable, mastery use)?

Figure 1

The Four Types of Test Scores Used In Educational Measurement



2. What are the effects of making the wrong decision? How much error should be allowed (dichotomous or continuous mastery)?

3. Should the test length be formulated for the individual or for the "typical" individual, or both (domain-referenced or mastery)?

4. Is the cutpoint or standard known, or was a process involving judgements used to set it (mastery)?

Answers to these questions indicate that there is no easy answer to the question of test length. Test developers can no longer assume "10-15 items per objective" (Popham, 1978) is adequate. Possible formulae which may be used to determine test length are plentiful but mathematically exacting. Novak and Lewis (1974), Millman (1973) and Hambleton (et al. 1978) present some of these equations. One of the best comprehensive reviews of formula for test length is that presented by Wilcox (1980) who concludes with two major problems in test lengths.

1. Most of the techniques developed are for use in dichotomous, or mastery, situations. Methods need to be established for situations in which competency-based tests are used as diagnostic or program evaluation tools.

2. The current level of mathematical expertise needed to establish test length is beyond that of most of the users and developers of competency tests. Methods need to be established which can be used by classroom teachers, educational evaluators and curriculum planners for both classroom tests and for certification decisions.

Reliability

The reliability or consistency of responses to competency-based tests and the decision based on those responses has been one of the major areas of concern and research since Popham and Husek's (1969) article. When developing or selecting a competency test, the user is faced with the need to show two types of reliability: 1) the consistency of the measurement and 2) the consistency of any decision based on measurement. Although not frequently acknowledged, norm-referenced tests also require both of these properties. In norm-referenced situations, decisions about an individual are often made after seeing how he or she stacks up against the group; in criterion-referenced situations decisions are made after seeing how the individual stacks up against the domain (Popham, 1978). In either case, reliability of measurement does not imply reliability of decision. It should be noted that reliability includes not only consistency over time (stability reliability) but also consistency over form (equivalent or parallel forms reliability). Good tests, whether norm or criterion referenced, should possess both stable and equivalent forms of reliability for an individual score and for any decisions made based on that score.

Using the schemata presented in Figure 1, Traub & Rawley (1980) provide test developers with a summary of the current state of the art in competency-based reliability. In the first situation, where continuous measures are used to make decisions about an individual, reliability

questions deal with the consistency of the decision or the losses that can be associated with incorrect decisions. There are two ways of describing the goodness of the decision made in this situation. First, one can check the correctness of classification. This is similar to test-retest reliability in norm-referenced situations where the squared correlation between administrations estimates the variability between observed and true scores. The second method examines the correspondence in decisions between parallel forms. This type of reliability is extremely important in domain-referenced situations where comparable scores must result from differing samples of the domain. In neither of these situations is the individual's true score known; thus probabilistic models have been developed to estimate the reliability. Since most competency measurements use continuous scores for decision purposes, the majority of the work done by theorists on reliability development have concentrated on this area. A review of these methods and the associated formulas may be found in Traub and Rowley (1980) & Hambleton (1981). The user should be aware that the mathematics involved are complex and, contrary to some norm-referenced forms of reliability, are not easily calculated.

In the second situation, measurement of a binary variable is used to make a binary decision, that is, the test score, which is either 0% or 100%, is used to make a decision. In this case, the reliability questions deal with the consistency of the decision and the consequences of errors. In this situation, correct classification can be distinguished from misclassification while in the previous setting, serious misclassification could have been distinguished from minor and no misclassifications. Several approaches have been developed for determining reliability in this setting. These may include the decision consistency approach (Swaminathen, et al 1974), the squared point bi serial correlation between the observed test score and the dichotomous decision (Harris, 1974) and models of forgetting and guessing probabilities (McReady & Dayton, 1977).

In the third setting, the variable being used is continuous and for descriptors purposes. The reliability concerns the measure. The purpose of the test is to determine an examinee's standing on the domain of tasks. Most reliability and generalizability theories developed for norm-referenced tests will perform adequately for this type of reliability.

For the fourth situation, a measurement, binary in nature, is used to determine standing within the domain. Reliability is concerned with the accuracy of classification or standing. Wilcox (1977a, 1977b) presents categorical methods of establishing consistency over time and form.

In examining the problems, and proposed solutions to establishing reliability for competency measurement, some conclusions can be drawn:

1. Reliability of decision making differs from the reliability of measurement or description. Both psychometric properties are needed for good measurement.

2. The goodness of a decision is best described by the use of expressing two probabilities: a) false positives--a mastery decision is made when it should not have been and b) false negatives--a non-mastery decision is made when it should not have been. Reliability measures

should always include estimates of both these errors. Many currently provide only the first.

3. "Most of the proposals encountered for evaluating the goodness of decisions involve relatively strong theory and require difficult computations. There are no provisions now available that the classroom teachers can be expected to use routinely.... If it is important for classroom teachers to be concerned about the goodness of decisions based on tests--and if teachers are to make their judgements in meaningful terms--then procedures must be developed that are simple to use and that provide reasonably accurate results" (Traub & Rowley, 1980, p. 541).

4. Competency-based measurements are increasingly used for two situations which the above types of reliability do not address--curriculum evaluation and surveys of achievement. In these situations, the focus in not on the individual in reference to a domain nor on decisions made about individuals. Instead, the focus is on group characteristics such as the proportion of a group responding correctly to a particular item or domain of items. The tests are not normative in that students are not compared--they are criterion in that they assess the achievement of mastery of the group on certain skills. The focus of reliability is on the adequacy of group description rather than individual description. The distribution of scores on these tests are typically skewed; thus traditional normative statistics are not appropriate. Advances in determining psychometric properties are needed in this area.

Validity

Test validity is the competency-based property that is most underdeveloped and underreported of all psychometric characteristics (Linn, 1980). Because competency-based tests have an increased focus on the content or domain, content validity has been held by some to be the most important type of validity needed for criterion-referenced measures (Linn, 1980). Content validity does not, however, provide sufficient means of indicating that the test is valid. Criterion and construct validity are also needed (Hambleton, 1980; Linn, 1970; Messick, 1975). A brief summary of each of these three types of validity follows.

Content validity for competency-based tests is derived from two considerations: 1) the definition of the domain and 2) the true representativeness of the items selected from the domain. In establishing content validity, the focus is on the items, not the behavior of an examinee. It is the representativeness of the test form, not the test score, that is being considered when content validity is established (Hambleton, 1980). To establish content validity, test developers must establish three properties: 1) item validity, 2) technical quality of items, and 3) representativeness (Hambleton, 1980). The first property involves the definition of the domain and the writing of items. Various methods have been developed by theorists to help establish the domain. Popham (1978), Hambleton, (1980), and others have delineated and/or reviewed various methods that can be used to establish item validity. The second property, technical quality, can best be established by a specialist in measurement theories. Establishing this property includes reviewing items for

inappropriate distractors, item bias, culture-free traits, guessing and risktaking probabilities, and test-wiseness. Empirical establishment may include inter-item correlations, discrimination indices and difficulty levels at both statistical and cognitive levels (Nunnally, 1978). The third property, representativeness, indicates that the test form presents an accurate representation of the domain. To achieve this, stratified random samples, grid checks, or Cronbach's duplication method (1978), may be used.

Criterion-related validity, including concurrent and predictive validity, is also necessary for competency-based tests. Many of the interpretations made from these tests involve the prediction of future performance. At the present time, only norm-referenced techniques are available for the establishment of criterion validity. Alternate techniques for competency tests need to be developed.

Construct validity for competency measures faces the same lack of development as does criterion validity. Because the term non-mastery involves a construct of attributes, attitudes and processes, some determination of the test's "decision validity" is also needed. This can begin with the establishment of the domain but must continue into examination of the validity of the placement of an examiner. That is, for construct validity, content validity is necessary but not sufficient. Possible methods of establishing construct validity may be used--Guttman scolograms analysis in dichotomous situations, factor analysis with continuous variables, and experimental methods involving the use of alternate domains (Hambleton, 1980).

In summary, the establishment of validity for competency-based measures faces at least four major issues:

1. The overuse of content validity as the only proof of the psychometric property. Test developers must remember, "Content validity is a test characteristic. It will not vary across different groups of examinees.... However, the validity of test score interpretation will vary from one situation to another" (Hambleton, et al. 1978, pp. 38-39).

2. In the establishment of content validity, more than domain specifications is required. Technical quality and representativeness are also needed. Both of the latter needs will require familiarity with advanced measurement techniques.

3. Both criterion and construct validity have been too frequently ignored in the establishment of competency measures. If the test is to replace a current measure or to predict future abilities, criterion validity must be established. If a decision is to be made based on the test score, or if the examinee is to be placed according to a score, construct validity must be present.

4. As was indicated with reliability, competency-based measures are also being frequently used for two other purposes: the comparison of a group of scores against a domain, and program evaluation. Once again, the validity for an individual score does not establish the validity of a group score. If competency-based

measures are to be used for these purposes, alternate means of establishing validity will have to be developed.

Setting Performance Standards

As indicated in Figure 1, one of the major purposes of competency-based tests is to provide data for decision-making. These decisions may range from the monitoring of a student in a classroom through promotion, certification and graduation. Hambleton (1980) indicates that "A standard is a point on a test score scale that is used to 'sort' examinees into two categories that reflect different levels of proficiency relative to a particular objective (or set of objectives) measured by a test" (pp. 99). In setting that standard, the test developer does not just set a value; there must be a rational, validated meaning behind the standard. The empirical and/or rational setting of standards has been one of the most researched and developed areas of criterion-referenced measures. Hambleton (1980, 1981) provides an excellent review of these methods. Many of the methods involve the use of collateral performance and involved mathematical or probabilistic formulae, thus limiting their useability at the classroom level. Even the arbitrary standard setting method (Popham, 1978) using techniques similar to a tribunal, requires extensive time and resources. Whatever method is used, a review of the techniques indicates at least four major issues:

1. Seriousness of an erroneous decision. How much harm will be done if a student is misclassified? One must consider not only the student's cognitive functioning at this point but also the social and developmental issues.

2. Need for collaborative performance data. The developer and test user must consider whether the test itself will be enough evidence for the final judgement or if other measures will also be needed. A secondary concern centers around the process to be used by examinees with disabilities that prohibit them from taking this exam. Should alternate performance measures be allowed in this situation? Are alternate methods as valid as the test? Who decides when each is applicable?

3. Replication of standards. What percent of items on the test is the equivalent of other methods of performance. What percent approximates real life performance? If a student can master a test, does that mean he/she can master 80% of the real life problems or 100%? Who establishes, and how, not only the relationship, and associated errors, but which are of lesser consequence; i.e., classroom tests vs. graduation or certification?

4. When using competency-based standards, how good is "good enough"? If results show that a student has mastered at 80%, has that student done a "good enough" job. Has the school done a "good enough" job? Is it "good enough" that 75% of the students master the criterion 80% of the time? In other words, how do decision makers use standards, established for determining if an individual has adequately mastered the domain, to determine if a group has adequately mastered a domain? Can measures evaluating how an individual stacks up against the domain be used to

determine how an individual stacks up against other students? Or how a teacher stacks up against other teachers? Or how a program stacks up against other programs?

Summary

In summation, it appears that competency-based measurements will have a long-term impact on the psychometric arena. At least 45 states now provide or require some form of competency-based measures, many school districts are already implementing them as part of the graduation procedure, and untold numbers of teachers are using domain-referenced tests in making classroom decisions. This widespread use of the measure has had an impact on the psychometric world. Not only are theorists rapidly attempting to develop the necessary underpinning of this mode, as is evidenced by even a cursory examination of the journals, but at the same time, test publishers, test developers and local educational testing staff are also creating and using tests. Whether this development and actual use of tests in real decisions is appropriate without an underlying theoretical foundation is a question of great concern. Whether the question will be answered is not the concern of this paper. The concern of this paper is the present lack of psychometric properties for instruments which are being used on a daily basis.

For each of the four psychometric properties presented above, specific implications or areas of concern have been noted. While there are other properties that should be examined and other concerns that could be noted, when looking at these summaries it appears that these concerns can be summarized into four major areas:

1. Mathematical competency. All of the psychometric properties discussed in this paper require extensive training in mathematics, for not only their development and understanding, but also their use. If competency-based tests are to be used in the school "next door" by "your" teacher, either extensive changes in training will be required or the formula will have to be developed to the point where alternate methods are available. For norm-referenced tests some progress has occurred, as is evidenced by the KR21 formula for reliability and the Spearman-Brown formula for test length. These formula can be used by classroom teachers and do not require extensive mathematical backgrounds. However, domain and mastery tests still need extensive development; in fact, current work indicates that formula may get more complex before becoming simple.

2. The comparative importance of psychometric properties for classroom measures and for graduation requirements. There appears to be a pervasive attitude that psychometric properties are not as important for classroom decisions or descriptions of a child as for certification or graduation decisions. Measurement experts might argue that it is better to establish properties for a series of small domains thus laying the groundwork for the establishment of one large domain.

3. Improper use of competency-based tests to determine group progress or program evaluation. As stated above, the purpose of a

competency-based test is to determine how an individual stacks up against the domain of skills. The current reliability and validity procedures provide probabilities of error for an individual decision, not for decisions about a group or program. Until a way is found to establish psychometric properties for such uses, the test should not be involved for either describing or making decisions about groups or programs.

4. The use of competency-based tests as measures of the affect. In general, competency measures have been directed toward the cognitive side of testing. Some attempts, however, have been made in the development of competency tests that measure affective variables (Popham, 1978). Several cautions should be noted about these attempts. First, the affective domain is extremely difficult to specify; the establishment of the domain specifications may require a lifetime of work. From a psychometric point of view this will necessitate some form of construct validity as well as content validity based on domain specification. Second, with only a few exceptions, most affective measures are future oriented; i.e., vocation, interest, etc. This will require the presence of criterion validity along with rigorously established specifications. Third, the establishment of a cut-off score; i.e., the "mastery" of an affect, is extremely difficult to establish. Fourth, there is a great tendency for test users to use the affective instrument as a measure of group progress or program evaluation. Again, different modes of reliability and validity would be required.

Where and how competency tests will be used in the future is not known. What is known is that they are being used now. The development of these instruments and the use made of the scores affect education on a daily basis. How much progress is made with, or damage done to, students and to education depends on the weight educators, students and the lay public place on scores that are, for now, frequently derived from instruments with little, or no, psychometric properties. It is the responsibility of persons involved in the use of these tests, or the decisions resulting from the scores, to inform others of the possible dangers involved. This responsibility rests not just on the shoulders of the psychometricians (whether developing the theory or the tests), but also on the shoulders of administrators, teachers, and others who are involved in the use of these instruments.

References

Berk, R. A. (Ed.) (1980). Criterion-referenced measurement: The state of the art. Baltimore, MD: The Johns Hopkins University Press.

Cronbach, L. J. (1971). Test Validation. In Thorndike, R. L. (Ed.) Educational Measurement 2nd ed. Washington, D.C.: American Council of Education, pp. 443-507.

Hambleton, R. K. (1981). Advances in criterion-referenced testing technology. In C. R. Reynolds & T. B. Gutkin (Eds.) Handbook of school psychology. New York: Wiley.

Hambleton, R. K. (1980). Test score validity and standard-setting methods. In Berk, R. A. (Ed.). Criterion-referenced measurement: The state of the art. Johns Hopkins University Press; Baltimore.

Hambleton, R. K., Swaminathan, H., Algina, J. & Conlson, D. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. Review of Educational Research, 48, 1-47.

Harris, C. W. (1974). Some technical characteristics of mastery tests. In C. W. Harris, M. C. Alkin & W. J. Popham (Eds.), Problems in criterion-referenced measurement (CSE Monograph Series in Evaluation No. 3). Los Angeles: University of California Center for the Study of Evaluation.

Linn, R. L. (1980). Issues of validity for criterion-referenced measures. Applied Psychological Measurement, 4, 547-561.

Linn, R. L. (1979). Issues of validity in measurement for competency-based programs. In M. S. Bunda & J. R. Sanders (Eds.), Practices and problems in competency-based education. Washington, D. C.: National Council on Measurement in Education.

Macready, G. B., & Dayton, C. M. (1977). The use of probablistic models in the assessment of mastery. Journal of Educational Statistics, 2, 99-120.

Millman, J. (1973). Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 43, 205-216.

Messick, S. (1979). Test validity and the ethics of assessment. Paper presented at the annual meeting of the American Psychological Association, New York.

Novich, M. R., & Lewis, C. (1974). Prescribing test length for criterion-referenced measurement. In C. W. Harris, M. G. Alkins, & W. J. Popham (Eds.), Problems in criterion-referenced measurement (CSE Monograph Series in Evaluation, No. 3). Los Angeles: University of California, Center for the Study of Evaluation.

Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw-Hill.

Perkins, M. R. (1982). Minimum Competency Testing: What? Why? Why Not? Educational Measurement: Issues and Practice, 1, 1-9.

Popham, W. J. (1978). Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. Journal of Educational Measurement, 6, 1-9.

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: A decision-theoretical formulation. Journal of Educational Measurement, 11, 263-267.

Traub, R. E., & Rowley, G. L. (1980). Reliability of test scores and decisions. Applied Psychological Measurement, 4, 517-545.

Wilcox, R. R. (1980). Determining the length of a criterion-referenced test. Applied Psychological Measurement, 4, 425-446.

Wilcox, R. R. (1977a). New methods for studying equivalence. In C. W. Harris, A. P. Pearlman, & R. R. Wilcox, (Eds.), Achievement test items - methods of study (CSE Monograph Series in Evaluation No. 6). Los Angeles: University of California, Center for the Study of Evaluation.

Wilcox, R. R. (1977c). New methods for studying equivalence. In C. W. Harris, A. P. Pearlman, & R. R. Wilcox, (Eds.), Achievement test items - Methods of study (CSE Monograph Series in Evaluation No. 6). Los Angeles: University of California, Center for the Study of Evaluation.

Florida's Experience in Implementing A Competency-Based Testing Program

By

Thomas H. Fisher

Introduction

Purpose

The purpose of this paper is to describe the implementation of a competency-based testing program in Florida's public school system. The task of designing and implementing the program was complex, and it will not be possible to explore in these few pages everything which occurred. For this reason, the paper will be limited to an overview of the most important aspects of the program design, its legal challenges, the results of the testing, and the lessons which have been learned.

Background

Florida has had an ever-increasing interest in implementing a quality educational program with a strong, centrally organized data base capable of describing various aspects of the educational enterprise. If one inspects the annual reports of the Commissioner of Education, required by law since the turn of the century, it becomes apparent that Florida has continually moved in these directions. In the early 1900's, the State was concerned about the mere existence of public schools in all parts of the State. This interest broadened over the years toward the provision of schools which protected the health and safety of the students (e.g., indoor plumbing, fire codes, etc.).

Beginning about 1930, the Commissioner's reports began to include statistical information which described the schools, the faculty, and the budgets. As data collection methods improved, the statistical information turned to questions of educational quality--first with the educational facilities and materials and then with whether or not students were learning what they should.

By 1971, Florida had introduced its first statewide accountability program with educational goals and a testing program, involving only a sample of students, designed to reveal the students' level of mastery of the 3 R's.

From 1971 to 1976, the Legislature modified the accountability program several times as it attempted to find the right combination of requirements and data gathering techniques. In 1976, the Legislature enacted a series of modifications which introduced the concept of holding students responsible for the attainment of certain educational objectives and holding the school districts responsible for the provision of an educational program which would lead to the attainment of these objectives. This was the beginning of the State's minimum competency testing program.

A Description of the Florida Competency Testing Program

The Basic Skills Component

The 1976 Educational Accountability Act (Laws 1976) required the State to define certain minimum student performance standards (MSPS) for grades three, five, eight, and eleven in the areas of reading, writing, and mathematics. These standards were to be adopted by the State Board of Education and used as the basis for the statewide assessment program.

The language of the act made it clear that it was the intention of the Legislature that students learn the minimum standards. However, the law did not specifically state that the assessment test which measured the standards must be passed before a student is promoted. This is a critical distinction between the Basic Skills component and the so-called "functional literacy" component described below.

In effect, the local school districts were responsible for teaching the required skills and students were supposed to learn the skills, but there was no absolute penalty for noncompliance. The only direct coercion implied in the law was the requirement that the State determine those schools which fell below a certain designated level of performance. These schools were to be identified as not meeting minimum requirements and efforts were to be made to improve the educational programs so that more students met the required minimums.

The minimum standards were simply broad performance objectives, each of which was further divided into sub-sets of more narrowly minimum skills (see Figure 1). The state assessment tests were directly keyed to the skills and were objective-mastery in design. Hence, there were multiple items measuring each skill with a mastery criterion established for both the attainment of the skill and attainment of the standard. Typically, students had to correctly answer eighty percent of the items in order to attain mastery status. The reports generated by the tests indicated how each student performed on each skill and standard. The reports were diagnostic in nature at the student level and were available in summary form at the classroom, school district, and state levels.

The tests were originally administered in October of each year with results available by late November. However, today, the eleventh grade test is given in March of the tenth grade to provide more time for additional instruction if necessary.

The "Functional Literacy" Component

The 1976 Educational Accountability Act also required students to pass a test of so-called "functional literacy" prior to graduation. This test was to measure the students' ability to apply basic skills to tasks of an everyday nature. The State identified twenty-four such skills which formed the basis of the first versions of the test (see Figure 2 for some sample skills).

There was a certain amount of controversy over use of the tern "functional literacy" because of the connotations of the term. The State Board of Education in 1978 changed the official designation of the test to the more

Figure 1

Examples of Standards and Skills

READING

Standard A: The Student will demonstrate knowledge of a basic vocabulary as determined by a specified word list.

1. Identify frequently used words by sight.
2. Identify the meanings of frequently used words presented in context.

WRITING

Standard D: The student will write for the purpose of supplying necessary information.

28. List information (who, what, when, and where) of a given message.
31. Include in any written message, invitation, or announcement, necessary information such as who, what, when, where, and how.

Standard G: The student will spell correctly.

49. Spell his/her own first and last name.
53. Spell his/her own complete address.
61. Apply generalizations for adding common suffixes.

MATHEMATICS

Standard A: The student will read and write numerals.

7. Read and write, in numerals, money values through five dollars.

Standard D: The student will determine equivalent forms of fractions, decimals, and percents.

25. Identify one-half, one-third, or one-fourth of a given region.
33. Identify a mixed number less than 100 that is equivalent to an improper fraction.

Figure 2

Examples of SSAT-II Skills

COMMUNICATION SKILLS

12. Determine the main idea stated in a paragraph.

15. Identify the cause or effect stated in a paragraph.

25. Obtain appropriate information from pictures, maps, or signs.

32. Include the necessary information when writing messages to make a request, to supply information, or to note an assignment.

47. Complete a money order or a check and its stub.

MATHEMATICS SKILLS

116. Solve real-world problems involving averages of no more than ten numbers and no more than two distinct operations.

120. Solve real-world problems involving addition or subtraction of proper fractions with unlike denominators of 2, 3, 4, 5, 6, 8, or 10.

132. Solve real-world problems by finding simple interest.

149. Read and determine relationships described by line graphs, circle graphs, or tables.

neutral "State Student Assessment Test, Part II" although the original language can still be found in the statutes.

The most important distinction between the "SSAT-II" and the "SSAT-I" (the basic skills component) was the fact that students were required to pass the former whereas teachers had only to "certify" student mastery of the basic skills for the latter. Students could not avoid passing the SSAT-II if they wanted a diploma from a Florida public high school. And, while five opportunities were available for students to take and pass the test prior to graduation, the idea was foreign to many educators who felt it was improper to require students to demonstrate any level of proficiency prior to graduation. It was this basic philosophical disagreement within the educational community which led to controversy, litigation, and a delay in implementation of the program.

The Results of the Program

When the first test results were released for the SSAT-II, 36% of the students failed the mathematics test and 8% failed the communications test. The failure rates among black minorities were much higher, hitting 77% in the mathematics test and 26% in the communications test. Educational observers and newspaper writers had a field day writing stories about the Florida experience. Time (1977) described the test results with the headline, "Florida Flunks." The Phi Delta Kappan devoted an entire issue to the topic of minimum competency testing and featured one article written with the title "Minimum Competency and Incompetence in Florida" (Glass, 1978) in which the author predicted that the graduation requirement would disappear within one year.

To the outside observer, education in Florida appeared to be in disarray and confusion after the test results were first announced. There was considerable debate, and attempts were made in various ways to delay the implementation of the high school graduation requirements. Several administrative challenges were filed against the State, and the State Board of Education devoted an unheard of nine hours to a debate over the establishment of the passing scores for the tests. However, the support of the citizens, the Legislature, and the news media did not lessen. Indeed, the "controversy" appeared to come more from out-of-state educators and testing specialists than from Floridians.

Eventually, the State won all of the administrative challenges to testing requirements, and attention was turned to more serious challenges posed in the Federal Courts in the case of Debra P. v. Turlington (1979). The State prevailed in the case but not without considerable effort.

The Debra P. case was a full-scale attack on the competency testing requirements in Florida. It was brought on behalf of ten black minority students who failed the SSAT-II on their first try. Attorneys for the plaintiffs attacked the basic concepts behind the law and the quality of the tests themselves. A key issue in the case was the fact that Florida's public schools had once been segregated. This was an important consideration because one could question whether or not black students had an adequate opportunity to prepare for the test. A second major issue was whether or not the teachers had actually taught the skills required for the test.

The trial judge ruled that the State could not implement the graduation requirement until sufficient time had passed for all students to have attended integrated schools. Past discriminatory effects had to be erased before new requirements could be imposed. The judge also stated that sufficient lead time must be permitted before the new requirements were imposed on any student regardless of the issue of past segregation of the schools. Unfortunately, the record is not entirely clear on what constitutes sufficient lead time.

Upon appeal, the Fifth Circuit Court of Appeals (Debra P. v. Turlington, 1981) ruled that the trial judge needed to hold further hearings to determine whether the State was teaching the students the required skills. The State thus embarked upon the design and implementation of an instructional validity study which apparently was the first ever conducted in any state.

The design and implementation of this study is more adequately described in Educational Measurement: Issues and Practice (1983) and will not be repeated herein. It is sufficient to say that the State collected a mass of information which was organized in a manner to convince the court that the students were being provided instruction in the required skills. The trial court issued an order up-holding the State's position and permitted the graduation requirements to be imposed with the class of 1983.

The imposition of the graduation requirement was met with general acceptance across the State. The teachers and students worked hard to meet the requirements, and were successful in doing so. As can be seen from Figure 3, the test results have increased over the years. Generally speaking, only about one to two percent of the students do not pass the test prior to their regularly scheduled graduation date.

An important element in this success story has been the State Compensatory Education Program. This Program provides over $30 million in financial aid to the school districts for assistance to students who have difficulty with the required competencies. Further, the Legislature has authorized full funding to any student who wishes to remain in school a thirteenth year, if necessary, to pass the required standards.

Another important factor in the story is the increase in achievement which has been brought by competition between schools and districts. Educators generally do not like to have their test scores published in the newspapers, but public access to the test results is very much a central theme in the Florida Accountability Program. This publicity combined with old-fashioned "score board" competition has contributed to considerable growth in student achievement where academic proficiency was not evident previously.

Lessons Learned from the Experience

Lesson #1: Be attentive to legal issues.

There is no doubt that the Florida experience has been characterized by considerable time spent in court. In a sense, this was inevitable as the Florida graduation requirements were the first in the nation. The principles were going to be challenged, and Florida just happened to be there.

Figure 3

Percent of Students Passing the SSAT-II on the First Try, by Ethnic Grouping

Communications Test

Test Administration Sequence

	1	2	3	4	5	6	7	8	9
Total	92	97	97	98	95	95	91	91	88
White	97	99	99	99	98	98	97	95	93
Black	74	89	91	94	88	89	89	80	74
Hispanic	93	98	98	98	94	93	91	85	80

Mathematics Test

Test Administration Sequence

	1	2	3	4	5	6	7	8	9
Total	64	74	78	78	78	78	78	87	84
White	76	83	85	86	87	86	86	92	90
Black	23	40	47	48	51	51	53	69	68
Hispanic	61	74	77	75	76	74	71	81	78

It is certainly true that we live in a litigious society. Citizens are not likely to sit quietly while the government imposes new standards and requirements which might be interpreted as arbitrary. Therefore, those who are moving toward competency-based educational programs would do well to prepare for the eventuality of litigation.

Excellent reviews of the legal principles related to competency testing can be found in several sources. Among them are Citron (1982 and 1983), McClung (1978 and 1979), and Tractenburg (1979). It is advisable for educators and testing experts involved in competency testing programs to be familiar with these issues and take steps to build programs which can withstand such legal challenges.

Lesson #2: Build tests of high quality.

If a competency program is to incorporate a testing requirement, attention must be paid to the quality of the tests to be used. In general, the tests must meet the requirements of the American Psychological Association's Standards for Educational and Psychological Tests (1974). The tests must be reliable, valid, and free of bias against any group of students. They must be administered fairly and consistently. The test results must not label or stigmatize students. Passing scores must be established with care and adopted by persons who have the legal authority to do so.

For a more complete discussion of the technical issues, one can refer to Fisher (1980 and 1983).

Lesson #3: Keep the public informed.

Competency programs generally are favored by the public, and are enacted by legislators responding to the demands of parents and citizens who are not satisfied with the quality of education being offered in the public schools. Based upon the Florida experience, it would seem that the competency program has provided a base upon which further educational improvements can be constructed. If the citizens believe that the educational community is trying to respond to the need for improved quality, then increased financial support for the schools is likely to follow.

However, to build such support, the public must be informed about what the new requirements are, what the tests measure, what the implications are if a test is not passed, what services are available to assist the students, and what the test results are for the students, schools, and districts. For this reason, in Florida, the news media is given complete access to the test results. And, to assist the reporters in interpreting the results, they are given the opportunity to actually take the tests under secure conditions.

Lesson #4: Use competition to raise achievement.

The State Board of Education in Florida set the stage for academic competition when it adopted the goal of making the State's educational system at least as good as that of the upper one-fourth of the other states. While measurement of the attainment of the goal necessarily will be open

to interpretation based upon available data, as a statement of intent it is flawless. Competition is at the heart of the goal statement.

In the same way, the competency based program in Florida has introduced the notion of academic competition similar to athletic competition. In athletics, the coach (i.e., the teacher) tells each person what his or her assignment is to be and how to accomplish the assignment. The athlete practices the assignment over and over, always trying to be better. Indeed, the athlete would never dream of coming to a game without practicing. Yet, in the academic arena, often students are not told what they are expected to learn; they may not practice sufficiently to perfect the skills; and more than one student has entered the competition arena (the test) without adequate preparation.

Last, the coach expects his team's results to be highly visible to the public. It is competition which creates the fans' enthusiasm for the game and pushes the teams toward greater effort. In Florida, the "score board" idea is being borrowed and applied to academic competition to encourage schools to better their past academic records or do better than a neighboring school. To illustrate, in one small rural district, test scores had always been dismal. The district had no traditions of academic excellence, and to many people, the very idea was preposterous. When the first test results were announced in 1977, the district scored at the very lowest among the sixty-seven school districts. The school teachers, administrators and students were determined to increase their rating. And, they succeeded. By 1984, the district was among the highest in the State, and the district took great pride in striving to be "first on the SSAT-II." Surely, the students benefited from the increased concern for their academic training. Mastery of the minimums will undoubtedly contribute to those students' ability to be more successful in school than they were before.

Conclusions

The State of Florida has embarked on an educational improvement program which is unique and timely. The program provides for the definition of certain skills which are important for students, the measurements of students' attainment of those skills, and public accountability for the schools and districts. The program has been challenged repeatedly in court and in public debates. Despite attempts to dismantle the program, it has survived and is successful.

This does not mean, however, that all other states or school districts will necessarily want to duplicate the Florida program. States and districts differ in their traditions, financing, and governance of education. What works in Florida may not be needed or desirable in other locales. However, there are some ideas coming from the Florida experience which are reasonable for others everywhere. Students should be told what they are expected to learn, and teachers should devote their effort to instruction toward those skills. Instructional activities should match the skills which are important, and the public should have access to information about how well the schools are achieving. These ideas do not seem to be unreasonable; indeed, they appear to be fair and equitable.

References

Citron, Christine H. Competency testing: emerging principles. Educational Measurement: Issues and Practice, 1982, 1, 10-11.

Citron, Christine H. Courts provide insight on content validity requirements. Educational Measurement: Issues and Practice, 1983, 2, 6-7.

Debra P. v. Turlington, 474 F. Supp. 244 (DC M.D. Fla. Tampa Div. 1979).

Debra P. v. Turlington, 644 F. 2d 397 (5th Cir. 1981).

Department of Education. Laws Related to Florida Public Education Enacted by the 1976 Legislature. Tallahassee, FL: author, 1976.

Fisher, Thomas H. Implementing an instructional validity study of the Florida high school graduation test. Educational Measurement: Issues and Practice, 1983, 2, 8-9.

Fisher, Thomas H. The courts and your minimum competency testing program--a guide to survival. NCME Measurement in Education, 1980, 11, 1.

Florida flunks. Time, December 12, 1977.

Glass, Gene V. Minimum competence and incompetence in Florida. Phi Delta Kappan, 1978, 59, 602-605.

McClung, Merle S. Are competency testing programs fair? Legal? Phi Delta Kappan, 1978, 59, 397-400.

McClung, Merle S. ECS Law and Education Center Footnotes. Denver, CO: Education Commission of the States, November 1979.

Standards for Educational and Psychological Tests. Washington, DC: American Psychological Association, 1974.

Tractenberg, P. L. Legal implications of minimum competency testing: Debra P. and beyond. Denver, CO: Education Commission of the States, October 15, 1979.

Conversation with the Speakers

Those Reponding To Questions During This Segment Included:

Dr. Charles R. (Bob) Davis, Moderator, Oklahoma State University
Dr. Thomas H. Fisher, Florida State Department of Education
Dr. Joseph R. Weaver, Oklahoma State Department of Education
Dr. Zoe Leimgruebler, Oklahoma State Department of Education
Dr. Daniel Selakovich, Oklahoma State University
Ms. Judy Leach, Oklahoma State Department of Education
Dr. Beverly L. Anderson, Education Commission of the States
Dr. David Yellin, Oklahoma State University
Dr. William E. Camp, Oklahoma State University
Dr. Kay S. Bull, Oklahoma State University

[Most of the questions were posed by unidentified members of the audience.]

QUESTION: What is the status at the current time in the state of Oklahoma as far as our mandated testing program? What is going to be done with the data? How is the data going to be broken down? How is it going to be disseminated?

ANSWER (Dr. Joseph R. Weaver): I feel pretty comfortable because I have some capable people here that I can refer to--or defer to--if I need some assistance. Judy Leach is here, from our Department; she's Associate Superintendent of Instruction, and she's been working with the Legislature in the process of writing the law (1466), which includes the student assessment program. Dr. Zoe Leimgruebler is in our Department and she has been working in the process of implementing that law and writing the RFP that has just recently been sent out to test publishers, and she will be involved in the administration of the program. So we can call on them for some specific answers.

I'll try to give you just a brief answer about the status. We just mentioned that the RFP has been written and has been sent out, and there is a timeline that has been established for the completion of the implementation of the program. I would like, if I could, to ask Zoe, to give us just briefly the key points in that timeline. She has been working on that for several months, and I think it would be good for her to share that with us.

ANSWER (Dr. Zoe Leimgruebler): We released the request for proposals on September 16th. The RFP is a legal document that goes out to all testing companies that might possibly have something that we could implement in Oklahoma. They are in the process of writing proposals at this time. We had a pre-bidders conference on September 27th--last Friday--and attendance at this meeting was mandatory for all testing company people that wanted to bid. We had eight companies represented, and among those eight

companies are those companies that produce the tests--the very tests--that the public schools in Oklahoma are using. None of those companies was left out.

The proposals are actually due to the State Board of Affairs by October 18th. Between October 18th and November 14th, we will use committees of teachers and administrators--but mostly classroom teachers--from throughout the state. In all, we will be having 85 people--teachers--involved in looking at the test materials, doing various reviews from technical factors all the way down to looking for the match of measurement of each suggested learner outcome at each grade level we're going to be testing. We have a technical advisory committee which will meet and look at the technical aspects--like reliability and validity and so forth. On the fourteenth of November, we will report to the State Board and the State Board will then make the selection.

Testing dates tentatively are set at March 18th through 20th, and March 25th through 27th. We have been asked a number of times: why those dates? We put two weeks in there so that all schools could avoid their Spring breaks. Anytime during March and April when we would have had it we would have hit somebody's Spring break. Also, we are having to back it up to be early enough so that the schools can get the results before school is out. We have many, many of our tiny rural districts that get out of school the first week in May. So that's why the dates are set for then. We tried to accommodate all 613 school districts.

QUESTION: Dr. Weaver, what is going to be done with the data?

ANSWER (Dr. Weaver): I am glad that Tom Fisher discussed this because I think that all of you are aware that the State Department of Education in Oklahoma is very sensitive to the issue related to disclosure of test results. We are sensitive to the concerns of school districts. However, you have to understand that when we do a state-wide testing program, we are going to have a bunch of data there, and there's no way that we can prevent that from being public knowledge. And so our intent is not to rank-order school districts in any way, but we are going to have to give some realistic information to the State Board of Education and to the Legislature. And what is done with the data once it leaves our hands and enters the hands of the State Board of Education and the State Legislature we may not be able to control. It is in the public domain then, and if the publisher of any newspaper in the state wants to take that information and do their own calculation and ranking, we may not be able to prevent that. And I'm just trying to tell you as honestly as I can what may happen. I think that's in a nutshell where we stand at the moment.

QUESTION (from Dr. Daniel Selakovich): Who is going to decide which content and which skills will be tested?

ANSWER (Dr. Weaver): Zoe talked about committees of people that represent the different subject areas that will be included in the testing, and they are the ones that are going to do it.

QUESTION (follow-up from Dr. Selakovich): How will they be selected? ... I know the legislation specifies the areas. But who is going to make the decision about the specific content that is going to be tested? In social studies, the law is more specific, for example, than in reading, math, and language arts, because it just mentions them. In social studies, it talks about--the law specifically mentions--American government, citizenship compared to other countries. Who is going to decide in social studies what the skills and the content are that are going to be tested?

ANSWER (Dr. Leimgruebler): I don't know if you know about our suggested learner outcomes. But we had committees of teachers from throughout the state who worked over the last two years to develop suggested learner outcomes. We're not calling them minimum competencies, but just exactly that--suggested learner outcomes. And the number-one criterion for selection of the standardized, norm-referenced test that we will use in the program will be how closely the test matches and measures those learner outcomes.

QUESTION (follow-up by Dr. Selakovich): How did you pick the teachers?

ANSWER (Dr. Leimgruebler): The teachers were picked based on their expertise at each subject level, at each grade level.

QUESTION (follow-up by Dr. Selakovich): How did you determine their expertise?

ANSWER (by Ms. Judy Leach): They were selected on the basis of nomination by other teachers, by administrators, by associations.

QUESTION: What kind of protection and what kind of recourse will the individual districts or the individual teachers have if this kind of information is used in the evaluation process or in a punitive way?

ANSWER (Dr. Weaver): The first thing I would like to say is this. What goes on at the district level is the business of the school district, and the State Department of Education has nothing to do with that. And we're not going to collect any data that deals with the performance of individual students and base the evaluation of teachers on student performance. If the local district wants to do that, and they have the means of doing it, that's their decision.

COMMENT (Dr. Thomas Fisher): I'm not sure that the question has an answer because the classic answer would be: if you are in a local bargaining unit which has a contract which specifies the manner in which teachers are to be evaluated, and if that contract is violated in some way--perhaps, as you imply, by using these test results as part of that evaluation--then obviously that is a contractual matter, and the state doesn't have anything to do with that. That is a matter between the local teacher, the bargaining unit, the grievance procedure, and everything else. So that matter is controlled strictly locally.

I think the question started out with some comments about ranking districts and schools and whether or not it was going to end up in the

newspapers and so forth, and the answer as I indicated from Florida's perspective is that that is one of the intents, one of the design factors for the program. The State Department of Education publishes a publication each year called "Comparative Ranking of Schools"--it's there. And if you want to know which one is the first district or the 67th district--it's there. And it is information available to the decision makers. In fact, the compensatory education law was written in such a way that it channels money to districts on the basis of their low scores on the state assessment. It's documented, educational need, and those people that have greater need get more money under that program.

COMMENT (Dr. Davis): "So if you score low, you get more dollars?"

ANSWER (Dr. Fisher): That's correct. Now, I might point out that it's great fun to come to educational meetings such as this and pound on the table about whether or not you are rewarding incompetence. I'll guarantee you that there is not a school superintendent anywhere that will stand up in front of his people and say: "We are going to do as poor as we can so that we can get more money."

COMMENT (Dr. Beverly Anderson): I think it's important to keep in mind what the purposes of the assessment are in determining these things, and I can just comment on what some of the other states have been doing. I know of one state where the state gives the results back only to the districts, and then the districts are the only way that newspaper reporters can get results, so that a reporter would have to go to every district and collect the information before he or she could make those comparisons. Also, there are some states that do sampling of students such that there's inadequate students from any given school or district to be able to make comparisons. So they totally wipe out the possibility of district comparison by that technique.

QUESTION: From your experience in Florida, would you speak to the timeline that was referred to here for Oklahoma for the selection of the test and the administration of it in the Spring of this year?

ANSWER (Dr. Fisher): Let me draw a parallel. I didn't mention it in my speech but we have a new testing program--yet another new testing program--in Florida: a norm-referenced testing program that's probably going to be very similar to what you're attempting to do now. We will be going to the State Board of Education this month for a policy clearance on the direction in which we are going to go. We will then write the invitation to bid, which will result in the same kind of identical screening process [which was described earlier for Oklahoma], involving 100 people or whatever, to select a test which will then be recommended to the State Board of Education for a multi-year contract. We don't see any point in doing it on the basis of a one-year contract because you don't want to change tests every year. So we'll be going for probably a five-year commitment once we make that decision to go with "Brand X." Now it takes three months in our law for the selection of a bidder and a winning contractor. So I anticipate that we will start this month; we will have a contractor lined up in April and enter into a contract; and we can test for the first time in the Spring of the following year.

Now that happens to be our timeline. That in no way suggests that your timeline is impossible. You have to do what you have to do. And if your law says you will have a test in place by May of 1986, you have to understand that the State Department of Education and the employees there are employed for the purpose of implementing the law. If you don't like the law, you go to your legislators to get redress. You don't go to the Department of Education and complain that you're doing the impossible. Departments of Education are quite frequently asked to do the impossible.

QUESTION: Is there equalization of financing in the state of Florida? Would that be affected by low performance?

ANSWER (Dr. Fisher): We are a state with only 67 county school districts and we have full state funding--everybody gets the same amount of money regardless of where you live--in terms of the basic funding formula. The compensatory education program provides additional monies on top of that.

QUESTION (follow-up for Dr. Fisher): Does the equalization of funds affect the performance of students?

ANSWER (Dr. Fisher): I don't know. I think it's a basic principle in my own value system that I think schools should be given the same level of resources. I say that because I lived in Michigan for ten years and resided in a very poor district which found it was absolutely impossible to raise adequate funds, and I taught in an adjoining district which was wealthy suburbia and they had no problem at all in raising twice--three times--as much money as the one in which I resided and sent my own children to school. So I'm a very firm advocate of full state funding programs, but when we talk about what makes kids do better or not--there's a thousand things that make kids do better or worse.

QUESTION: I see a threat if the results are published of those students and the "uneducated public" runs their kids up to the next district ten miles up the street that has better schools, and as a result the smaller rural districts have big financial problems. Do you see that as a threat?

ANSWER (Dr. Fisher): There are those people who believe in tuition vouchers right now as a system of bringing competition into the public schools to force greater effort and greater quality. I will not take a position on that issue. But if you're watching the current literature you will see quite frequently debates from Congress and so forth about that particular issue. My own belief is that competition is very much a way of our life, and competition gives people a sense of urgency to move forward. It shakes us out of our complacency. I really believe that I work for the people of Florida. I work for the Legislature. The Legislature states what the policy will be for education in the state of Florida, and it's my job to implement the policy. It's not my job to make policy. The people, it seems to me, own the schools.

Someone gave me a piece of paper in Florida yesterday which I was reading as I came up on the airplane, and in this note this person referred to the

Legislature as my enemy. And I was offended by that. I don't view the Legislature as being my enemy, and I don't view public disclosure of test results as being negative. I think the schools belong to the people; the people have the right to that information. I think they should have that information. They deserve it. And if we don't like the results because it makes us squirm out of our complacency, I don't see that as being bad.

COMMENT (Dr. Selakovich): I would like to comment on this easy generalization about the relationship between input and outcomes. Dr. Fisher has elected to indicate that there's a relationship. And of course anybody who has read the research from Coleman to Jencks and anybody since knows there's a clear indication that there is not a relationship between dollar input and educational output, which really raises the whole question of the viability of the testing program, particularly in a state like Oklahoma, where the major thrust of the assessment program is to identify people who need remedial treatment. Any sociologist, any student who takes Sociology I at OSU (which is a very good course), would understand that there are variables over which the school has no control--but there is a relationship between social class and achievement, which is a problem that's pretty much ignored by the whole issue of state assessment.

COMMENT (Dr. David Yellin): I would like to perhaps question this notion that competition among schools and competition among districts is going to benefit us all and in particular the children of this state. I could envision a scenario as follows: where districts would be so concerned about scoring high on the test that they would draw the wagons in a circle around their resources, around their specialists, around all of the materials that they have for schooling, and would be unwilling to share them with a neighboring district. And I think if that becomes the case, the overall quality of education for the majority of students may decline rather than increase. This notion that competition is going to make us all work harder--it may in effect eliminate what little cooperation we have among schools within districts and between districts. I think that we should consider the value of cooperation among educators as being as, if not more, important than competition.

COMMENT (Dr. Fisher): I heard it once said that the only problem with the wagons in the circle analogy is that in education we quite frequently circle up the wagons and shoot inward.

COMMENT (Dr. William Camp): I think another aspect of this that should be addressed very quickly is the fact that, as I talk with administrators, I'm finding that there is a concern because there's not enough information going out to administrators, to schools, to the public about what is expected and what is needed from the program. And so, I just would like to make the point that it's going to be a very good year for communication--communication from several aspects: from the State Department and Legislature, administration, teachers, students, and public. But even more important: with the emphasis that we are putting on testing, we need to point out to our public that there are a lot of other things that schools do extremely well that will not be measured by this testing program, and emphasize these too.

QUESTION: I keep hearing these terms: "release to the press," "how do we rank"? Are we really doing this to test what our students need to know, or are we doing it for a public relations ploy for the state Legislature, to assure them we will test everybody, we will give it to the press, and all of our patrons will pick up the paper and say: "Boy, our kids do good" or "Our kids do bad"? It seems like we're forgetting what we're doing. We're trying to assess what our students know so we can educate them better. We're not doing this for a public relations ploy, and that's what it sounds like everybody is leading to, and I would like the panel to address that. What is the purpose of it? Are we testing or are we using it as a public relations tool?

ANSWER (Dr. Kay Bull): As far as I know, after having studied educational philosophy and pedagogy for the last--at least--15 years, there is no educational rationale for a state-mandated assessment program. There are plenty of political and plenty of economic rationales, but it doesn't have to do with doing better for kids, because it doesn't deal with the issues of how the individual kid is performing in relation to other children, which is what we need to be working with. It has to do with how do we assign dollars, how do we administrate, how do we legislate. It doesn't have anything to do with the education of children.

COMMENT (Dr. Camp): Looking at a number of the responses coming out--from the request for proposals and other things--the goals of the program appear to be directed towards assessment and individual use within districts. The material will go back to districts at the same time as it goes to the State Department. If it is used this way, then it should be an excellent program. And I think this is where, as administrators and teachers, State Department, we all need to work together.

QUESTION: Do you feel, then, that the test that is going to be developed is going to be applicable in all 473 districts? Unlike Florida that has 67 school districts, we have 473 independent districts that range all the way from Panhandle farming communities to the inner-city schools. From the tests that I've been around, you have to take into consideration ___ (not decipherable from transcript) what Dr. Selakovich was referring to earlier, so that ___ (not decipherable from transcript) are not going to be played into this.

ANSWER (Dr. Camp): Hopefully, it will be only a part of a district's testing program. It's the only answer you can give to that, because it will not apply completely to all districts, but it will be just one more set of data available for the district and for those teachers to use.

COMMENT (Dr. Bob Davis): What Dr. Camp is saying is that this is just a single item that will be a portion of the procedure utilized within the particular district. I think Senator Shedrick said this is not the only way we expect you to assess the students in your district, that we expect you to spend dollars there as well. Many districts already have. I know working as a school psychologist in a suburb school of Oklahoma City, I talked to Bill Hodges, who was superintendent, as to the results of our testing, and we made decisions as to the emphasis within the school district based on that, in part.

COMMENT (Dr. Camp): Bob, I think it's very, very important that we tell our publics this year what we are doing well. In other words, we've got a lot of emphases, there are going to be comparisons. In schools in Oklahoma, we have people, as I said before, competing for our teachers, from outside states--we have good educational programs. We need to tell people what we are doing well. I think it's very important this year and in the next few years to continue.

QUESTION: For schools that are participating in Chapter I programs, we already have assessment programs in place. Are you advocating that these tests that are mandated by the state are going to be in addition to those that we already are using at every grade level? Could we drop the testing that we are doing? Or should we be testing these children in March and then again in April?

ANSWER (Dr. Weaver): That is a decision that you are going to have to make in the local district. You'll need to take a look at what the state testing program looks like and then determine if you can fit that into your local testing plan. And if you need a test in addition to that, then I think you'll have to make that decision for yourself.

QUESTION (follow-up from same questioner): Well, isn't it almost required that we drop the testing that we have been doing, the norm-referenced testing, at 3rd, 7th, and 10th grades, because of the time that it would take to do this testing and the time that it would take to do the testing we are already using?

ANSWER (Dr. Weaver): I'd say that I agree with you. That is something you will have to consider because there would be considerable time involved.

QUESTION (follow-up): How long will the test require?

ANSWER (Dr. Leimgruebler): We are not developing a test. The Legislature has called for us to administer a standardized, norm-referenced, off-the-shelf test just like you normally administer every year in your schools anyway. So you know about the time it takes--somewhere around 7-8 hours, spread over a three-day period.

QUESTION: Did the Legislature ask school districts to see how many school districts were already testing with nationally normed tests before they enacted this, or did they just decide they were going to do it?

ANSWER (Dr. Weaver): I don't think the Legislature took a survey, but there was a survey done by our Department last year and that information was provided for the Legislature in their decision-making process.

QUESTION: I'd like to go back to the timeline just a little bit. I noticed from the murmur in the audience a while ago when they were talking

about the people who helped to do the suggested learner outcomes that a lot of us didn't know anything about it. And I'm wondering when they get 85 people together, from teachers and principals, in November to go over the test items to see if the test is going to meet those learner outcomes, how many of the districts are going to be involved--if we've got over 400 and there are only 85 people on those committees?

ANSWER (Dr. Weaver): You understand that it would be impossible to get 400+ or 600+ school districts involved in working on a committee. So what you have to do is put your faith in the sample that's used to do this work. And you have to understand that the people that are working on these committees are representative of all of the school districts in the state, and the only thing I can tell you is that this process has been used before. We used the same kind of process in developing our teacher certification test--we used representatives of the universities, of the public schools in the state, and the same kind of representation will be involved in the committee work on these tests.

QUESTION: I'm just curious about how many people in here were contacted about this? I would like to know more about that--on the selection. I think that maybe we should be divided up in districts and have a little bit more local input. Because I think that that is a fear that everybody has--that all of sudden they are going to get something sprung on them and that it's a few schools that decide for the whole bunch of us. Now I don't know how many in here actually participated in it. I have a second question for the man from Florida, who had a lot of guts to be here: I believe in your testing theory as such, we're just a little bit scared of the content. We'd like to have a little bit more input, and I think that's a general concern. The question I have concerns special education students: do you figure those students, like learning disabled students, in with the rest of them, and if so, if you have one particular district that has many of these, wouldn't that encourage them not to go to school or to move elsewhere to other districts?

ANSWER (Dr. Davis): In the state of Oklahoma, the law specifically states that students covered under Public Law 94-142, that is, with an individualized education program, will not be included in the testing program.

COMMENT (Dr. Selakovich): Well, the first crack out of the box, the first possible lawsuit, will be some parent who says you're denying my child the opportunity to be measured with the rest of them.

ANSWER (Dr. Fisher): In Florida we do have special arrangements for exceptional students. I might point out before describing those, however, that there was litigation in Illinois last year over the very issue that you're talking about--in terms of whether or not exceptional education students could be deprived, or denied, a diploma based upon a norm-referenced test which is used in a minimum competency testing program. The Courts held that the school district could, in fact, do that. So, before you get all excited about it, you might want to check out that court case.

Now, in terms of what Florida does, we have minimum student performance standards for the regular students; we have minimum student performance standards for hearing-impaired students; we have standards for EMH students; we have standards for TMH students. We have a state hearing-impaired test; we have a state EMH test. Students are expected to meet their own standards. But if you want a regular diploma, then you must demonstrate that you have met all of the criteria for a regular diploma--that means you have to take world history, you have to take nineteen credits, you have to pass the test--whatever the case may be. We do not deny anyone access to the test. But just to be fair in terms of the public reporting, we do remove from the publicly reported scores, the scores of exceptional education students who may choose to take the test just to see if they can get past the hurdle. We also have braille versions of the test, large-print editions of the test, and everything else you can imagine in terms of special adaptations.

COMMENT (Dr. Anderson): In all this discussion, I think it's real important that we remember how different states are in terms of who determines curriculum. Now Florida is probably at the extreme in terms of (and Tom may disagree with me here) the state having a lot of control over defining the curriculum; they totally support the schools, and all of that. And there are other states, for example, Colorado, where its constitutionally illegal for the state to define the curriculum. And so, we have that whole spectrum in there, and I think that as we look at testing, we need to keep in mind that it has become a very real tool for defining the curriculum in a way and that people don't recognize that that is what it is doing. Or at least it is not put out front that that is happening. So I think that as we debate the role of state testing we're really largely needing to be thinking about the role of who is defining curriculum--state and local--and what that combination of roles is in any given state. And it varies widely.

QUESTION (as repeated by the Moderator): What are the guidelines as far as who is going to administer the test and the provisions for security of the test?

ANSWER (Dr. Leimgruebler): The regular classroom teachers who normally administer a standardized, norm-referenced test each year in your testing program--your classroom teachers will be the ones administering the test. The only way we know to handle security is to have the testing company send those tests directly to your school buildings and then have them mailed directly back to the test company after the test is over. We will need help locally, with one person in each building taking the responsibility of getting those test booklets distributed correctly, and so forth, and then when it's over, checking the materials before they are mailed back.

IN RESPONSE TO A REQUEST BY A MEMBER OF THE AUDIENCE, THE MODERATOR ASKED FOR A SHOW OF HANDS OF PEOPLE IN THE AUDIENCE WHOSE DISTRICTS WERE

INVOLVED IN THE DEVELOPMENT OF THE SUGGESTED LEARNER OUTCOMES OR WHO WERE TO BE INVOLVED IN THE SELECTION OF A TEST. IN RESPONSE TO THIS, DR. WEAVER ASKED TO MAKE ONE LAST COMMENT:

COMMENT (Dr. Weaver): Test development can not involve a process where everyone in the business of public education is involved--except through representation. We can't have a vote on each item that goes into the composition of the test (unfortunately--maybe that would be good if we did that). The second thing is: I don't think you can draw any conclusions about the representativeness of the committee that worked on, or will work on, this process by a show of hands of the people who are here today. You people are here voluntarily, at the invitation of Oklahoma State University, and in no way are you representative of the total school education population in the state of Oklahoma, and so I'm trying to draw an analogy here.

State-Level Involvement in the Competency Testing of Students

by

Carolyn W. Smith

I am speaking to you today from the perspective of a state education policy-maker. I want to begin by making sure that you understand how I view the function of the state and of state policy in the educational process.

The function of the state and of state education policy, first of all, is to create a context in which the best possible education can occur. It is the state's responsibility to enable education practitioners to do their jobs by providing them with the necessary resources--textbooks, curriculum materials, libraries, computers and other equipment, facilities, information, and opportunities for professional growth. The state provides these resources in a variety of ways ranging from funding basic operations to running innovative pilot programs and workshops to generate and disseminate information.

The second function of state education policy, or at least of state policymakers, is to account to the state taxpayers that their dollars are being spent in the most effective and efficient manner. This is true even in the best economic circumstances but is even more true where private and public resources are scarce. State government is the agent which must assure the state taxpayers in Boise City that their state tax dollars are being spent in the most educationally effective and efficient manner in Poteau. And vice versa. The state does this by establishing minimum standards and guidelines, and by assessing the results or output of the educational process.

I believe we have made great strides in improving education policy and practice in Oklahoma in the past five years. Our overriding commitment, especially during this past year, has been to enhance the critical role education plays in Oklahoma's economic development and overall quality of life. We have recognized that public education in Oklahoma must prepare our citizens to function and prosper in a world that is technological, international, and constantly changing. We have therefore sought to establish state policies directed toward improving instructional productivity and quality assurance.

In order to achieve this goal, we have, over the past five years, implemented policies designed to elevate the teaching profession: strengthen curriculum, standards, and learning resources; improve the effective use of class time; and make better use of available resources. We have implemented policies designed to elevate the teaching profession because we have believed that, in order to improve educational productivity, our first priority must be to assure that we have the most qualified people teaching our children. We must, therefore, provide screens to screen out the less capable and less dedicated active and prospective teachers; we must provide magnets to make the teaching profession more attractive in order to retain the many high-calibre active teachers we have and to attract highly qualified students into teaching; and we must provide oppor-

tunities and requirements to assure that teachers have the knowledge and skill necessary to be effective in the classroom. In order to make the teaching profession more attractive, we have increased teachers' basic salaries by 12%, increased retirement benefits, and we have created special scholarships and forgivable loans for prospective teachers.

However, money may be a necessary condition of an attractive teaching profession but it is not a sufficient condition. In order to make teaching more attractive, we have also sought to strengthen its professionalism. We have increased our expectations of current and prospective teachers, certainly to screen out the less capable and less dedicated, but also to attract and retain the more qualified. Higher standards of performance and professionalism do attract the kind of people we want teaching our children; the people who welcome a challenge and are willing to make the most strenuous effort on behalf of their personal growth and the betterment of society. In 1980, we established a model staff development program and during this past legislative session we established professional training institutes and provided tax incentives for businesses providing summer employment for secondary math and science teachers and for those releasing employees to teach part-time. These initiatives are intended to provide active teachers with opportunities (resource) as well as requirements (accountability) for professional growth; that is, for a continual upgrading of the skills and knowledge they need to be effective as teachers in a changing world. We have also, during this past session, strengthened the required evaluation of teaching performance, again with the intention of providing information for good teachers which will help them to be better teachers (resource) and also of assuring the public that we have qualified people teaching our children (accountability).

We have taken the same approach with prospective teachers. In order to attract our best students, and assure that they have the skills and knowledge necessary to make them effective teachers, we are currently developing an educational college admissions test which will complement the strengthened teacher preparation programs, certification requirements and entry-year assistance program which we initiated in 1980. All of these efforts provide prospective teachers with invaluable guidance and information on their own strengths and weaknesses which they can use to improve themselves as teachers. They also provide the public with the necessary assurances that state tax money is being invested in high quality teaching personnel.

Our efforts to improve the effective use of class-time have included the state board's ten-day absence policy and the phased class size reduction in grades 1-3. These efforts have also been motivated by the double responsibility to provide education practitioners with adequate resources (time and energy), and to assure the state taxpayers that the money they are spending to keep school buildings open to keep qualified personnel in the classroom is being used in the most efficient and most educationally effective manner. Reducing class-size in grades 1-3 might increase personnel and facility costs, but it will increase the educational return on the public's initial investment by far, far more than the increased cost.

Many of our efforts to improve curriculum, standards and learning resources have also been guided by this principle--that sometimes you have to invest more to greatly improve the productivity of your initial investment.

No matter how many children spend how many years in school, they will not come out prepared for today's world if they do not have access to updated textbooks, computers, and adequate library resources. Therefore, during this past session we more than doubled state support in these areas. In addition to these "heroic" efforts to provide adequate resources, we also established new educational services for blind and visually handicapped children; we strengthened programs for gifted and talented children; and we increased basic support to schools by $30 million.

But providing adequate resources to education practitioners involves more than just adding on. State policy-makers must, because of their responsibility to the taxpaying public, assure that available resources are being distributed equitably; that is, that they are going where they are most needed. This is why, in 1981, we revised our state equalization formula, which is the mechanism through which basic state aid to common schools is distributed. We revised it to assure that state aid goes where the need is, where the kids are and/or where local resources are not.

And we further strengthened the formula during this past session by distributing all salary increase funding through the formula. State taxpayers are not getting a fair return on their investment if there are children in any part of the state who are not being educated because of a lack of resources. We all lose the benefit of their productivity; we all pay the price of their lack of productivity; and they lose their birthright as American citizens to personal growth and economic opportunity.

State policy-makers must also assure that available resources are pooled where necessary and educationally appropriate. Therefore, we established incentive grants to encourage school districts to set up cooperative programs in subjects where equipment or facility costs may be prohibitively high or where qualified teaching personnel may be in short supply. We also authorized a centralized state math/science high school to assure that highly capable students in the smaller, poorer districts have the same access to expensive but educationally necessary programs as students in the larger, wealthier districts.

Where does the statewide and state-funded student testing program, which we initialized during this past session, fit into all of this? By now, that is obvious to many of you. From the perspective of a state policy-maker, the statewide and state-funded student testing program is another element of our effort to strengthen curriculum, standards and learning resources. We view it, first of all, as an informational resource which we are providing students, parents, teachers, administrators and school boards. The testing program should provide information on students' individual strengths and weaknesses, which students, parents and teachers can use to assure that education programs are meeting the students' needs. The testing program should provide information to teachers that will enable them to improve their instruction; it should provide information to administrators and school boards, which, when incorporated into their annual curriculum evaluation and school improvement plan review, will help them to design more educationally effective programs.

The student testing program will also assist state policy-makers in their role as agents of the public's interest. It will provide us with information on the effectiveness of our educational system and programs at pro-

ducing productive, informed and fulfilled citizens. We will be better able to evaluate policies we have already implemented and to fashion new policies that will improve the likelihood of better education occurring. And it will provide the public with another kind of assurance that its money is being well-spent.

I recognize the limits on the usefulness of quantitative measures, or even of qualitative measures, of educational output in evaluating the effectiveness of educational programs and policies. But to say that their usefulness is limited is not to say that they are useless. Along with many other resources and assurances I have discussed today, I believe that the student testing program will help all of us to continue improving the quality of education we offer our citizens. By doing so, it will make a significant contribution to the future of our state.